A Modular System for Site-Specific Modification of Native Oligonucleotides

---

Dissertation

zur

Erlangung der naturwissenschaftlichen Doktorwürde

(Dr. sc. nat.)

vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät

der Universität Zürich

von

Igor Oleinich

aus der Ukraine

Promotionskomitee

Prof. Dr. Eva Freisinger (Vorsitz und Leitung der Dissertation)

Prof. Dr. Roland K. O. Sigel

Prof. Dr. Nathan Luedtke

Zürich, 2013

Die vorliegende Arbeit wurde von der Mathematisch-naturwissenschaftlichen Fakultät der Universität Zürich im Herbstsemester 2013 als Dissertation angenommen.

Promotionskomitee: Prof. Dr. Eva Freisinger (Vorsitz und Leitung der Dissertation)

Prof. Dr. Roland K. O. Sigel (Chair)

Prof. Dr. Nathan Luedtke (Chair)

Acknowledgements

To educate somebody means to share information and to give chances. I thank Prof. Dr. Eva Freisinger for sharing with me her knowledge, her patience and giving me the opportunity to do first step in science.

My special thanks to the members of my PhD committee, Prof. Dr. Roland K. O. Sigel and Prof. Dr. Nathan Luedtke for kindly acting as referee.

Further thank goes to the whole MT group, Prof. Dr. Milan Vašák, Dr. Augusto César dos Santos Cabral, Dr. Barbara Leitenmaier, Dr. Silke Johannsen, Katsiaryna Tarasava, Jovana Jakovleska for being me great labmates, and also our guests visiting shortly or longer the lab, Aleksandar Salim Ammon, Alessia Dürst, Blanca Puigdelloses, Mansoore Esmaili, Manuela Bruetsch, Matthias Winnerlein, Michelle Burkart, Philipp Pfingstag, and Sonja Giger.

Special thanks go to Dr. Jens Löbus for his help and support inside and outside of the lab.

The special thanks also go to David Egloff, the best colleague in the world. Without his comments, knowledge and perfect example of science work this thesis was not able to be done.

My acknowledgement goes to Prof. Dr. Gilles Gasser for his advices and sharing experience in work with artificial nucleic bases. In the same time, I wish to thanks Anna Leonidova for help in solid-phase synthesis of peptide nucleic acids and, that is the most important and time consuming, in their following purifications.

I am also grateful to Dr. Serge Chesnov for his friendship, support and, above all, for making my way to this work possible *via* nice MS spectra.

I thanks to Prof. Bernhard Spingler for his support in the beginning of work.

My special thanks to all those people who made this thesis possible, also the ones which are not specially mentioned here. Without your help and support, I could not have accomplished this challenging work.

I also like to thank for the great financial support from the Swiss National Science Foundation, the Swiss Academy of Science, the Swiss Chemical Society, the University of Zurich, the Inorganic Chemical Institute, and the CMSZH graduate school.

Finally I thank my family, who always believed in my dreams.

Abstract

Functionalization of native ribonucleic and deoxyribonucleic acids is of paramount importance in all fields of research in life sciences. The most frequently applied manipulations include formation of strands breaks, interstrand cross-links or the covalent attachment of new groups and functionalities. A variety of bioorthogonal groups and modifications are in use, e.g. for therapeutic delivery and targeted release of drugs, as well as identification, sequencing and imaging of RNA. Surprisingly, despite the great demand for modified oligonucleotides, only two strategies dominate the chemical manipulation of nucleic acids. The most widely used strategy relies on the synthesis of modified nucleotides and their incorporation into the oligonucleotide sequence during solid-phase synthesis. Unfortunately, this strategy is limited by the size of the modified oligonucleotides that is usually significantly shorter than 100 base-pairs. The second strategy involves the treatment of native oligonucleotides with small organic compounds resulting in non-selective and multiple modifications of the target sequence. Due to the limitations of the aforementioned strategies, the development of a method for site-selective introduction of functional groups into long native DNA as well as RNA sequences would present a highly desirable tool for various future applications.

In this work we propose a strategy for site-specific modification of long oligonucleotide sequences. The strategy is based on a new modular system consisting of three parts: a short (<25 bases) oligonucleotide recognition sequence, a reactive group that is specifically designed to generate the respective modification, as well as a linker between recognition sequence and reactive group that often also contains a cleavage site. The reactive group should be chemoselective, react only with one nucleobase type in the presence of the three others and allow modification under mild conditions to preserve the stability of the oligonucleotide strand. The modification protocol consists generally of two steps. In the first step annealing of the modular system, also denoted the short recognition sequence (SRS), to the complementary region of the target single or double stranded oligonucleotide *via* Watson-Crick or Hoogsteen hydrogen bonds takes place. Formation of such a rigid duplex or triplex DNA structure localizes the respective reactive group close in space to the target nucleobase that results in annihilation of steric hindrance so that in the second step reaction between reactive group and nucleobase can occur.

In chapter 3 we demonstrate this approach for site-specific modification of the cytosine-C4 position catalysed by sodium hydrogensulfite. By varying the length of the linker and including a hydrazide-derivatives as a reactive group we increase the efficiency of cross-link formation up to 60%. In chapters 4 and 5 we used a bromoacetate-derivative as well as an acrylamide moiety for

site-specific modification of guanine and thymine target nucleobases. As a proof of principle, final interstrand cross-linked products were analysed by MALDI-TOF MS and enzymatic digestion with trapping of the covalently linked dinucleotides.

Incorporation of a cleavage site between the reactive group and the recognition sequence allows us to chemoselectively cleave the interstrand cross-link to produce a single oligonucleotide strand with a specific modification. In chapter 3 we demonstrate the possibility for oxidizing a cleavable diol-containing linker resulting in the formation of a highly reactive aldehyde group and their subsequent reductive amination to result in an amino-fluoresceine derivative. In addition we also used cleavable disulfide linkers between recognition and target DNA strands in chapter 4 and 5.

Finally, in chapter 6 we show a general approach for the site-specific labeling with the modular system for fluorescence applications. A 50-mer target DNA sequence that contains 14 guanines was site-specifically modified with a free thiol group. This highly reactive thiol group was labeled with a maleimido-sulfo-Cy3 dye and formation of the final product was proved by MALDI-TOF, RP-HPLC and PAGE. In the key experiment, the Cy3-modified strand was annealed to a complementary Cy5-acceptor strand and we observed increased emission of the acceptor dye most likely due to energy transfer between the two dyes.

Zusammenfassung

Die Funktionalisierung von natürlichen RNS- und DNS-Strängen ist für die naturwissenschaftliche Forschung von grosser Bedeutung. Die am häufigsten angewandten Modifizierungen beinhalten die Durchführung von Strangbrüchen, die Verbrückung zweier Stränge und den Einbau neuer funktioneller Gruppen. Eine Vielzahl von verschiedenen bioorthogonalen Modifizierungen wird heutzutage für therapeutische Zwecke sowie zur Identifizierung, Sequenzierung und Visualisierung (Imaging) von RNS-Molekülen verwendet. Trotz der grossen Nachfrage nach derart modifizierten Oligonukleotiden werden heute hauptsächlich lediglich zwei verschiedene Synthesestrategien angewandet. Die am weitesten verbreitete Methode ist die Herstellung von modifizierten Oligonukleotiden mittels Festphasensynthese. Allerdings ist diese Methode auf kurze Stränge mit bis zu 100 Basen limitiert. In der zweiten häufig verwendeten Methode werden natürliche Oligonukleotide mit kleinen organischen Molekülen behandelt. Dies führt jedoch in den meisten Fällen zur Anhäufung von unspezifischen Modifikationen in den Zielsequenzen. Aufgrund dieser Einschränkungen ist die Entwicklung einer neuen Methode zur punktspezifischen Einführung funktioneller Gruppen in längere natürliche DNS- und RNS-Sequenzen ein dringendes Bedürfnis für zukünftige Anwendungen.

In dieser Arbeit wird eine neue Methode zur punktspezifischen Veränderung von Nukleinsäuren vorgestellt, welche auf einem dreiteiligen modularen System basiert. Die Bausteine dieses Systems sind ein kurzes (<25 Basen) Oligonukleotid als Erkennungssequenz, eine reaktive Gruppe, welche jeweils spezifisch für die gewünschte Modifikation entwickelt wird, sowie ein Molekülbaustein, ein sogenannter Linker, welcher die Erkennungssequenz und die jeweilige reaktive Gruppe miteinander verbindet. Der Linker kann des Weiteren eine spaltbare funktionelle Gruppe enthalten. Die reaktive Gruppe sollte chemoselektiv sein und mit nur einer der vier möglichen Nukleobasen reagieren. Zudem sollte die Reaktion unter Bedingungen ablaufen, in denen das Oligonukleotid selbst stabil ist. Die Modifizierungsreaktion läuft üblicherweise in zwei Schritten ab. Zunächst findet eine Hybridisierung des modularen Systems, welches auch als reaktive Kurzsequenz (Englisch: *short recognition sequenc*e (SRS)) bezeichnet wird, mit dem komplementären Teilstück einer einzel- und doppelsträngigen Zielsequenz über Watson-Crick oder Hoogsten Basenpaarung statt. Durch die resultierenden stabilen Duplex- oder Triplexstrukturen wird die reaktive Gruppe in unmittelbarer Nähe ihrer Zielnukleobase positioniert, so dass im zweiten Schritt die Reaktion zwischen reaktiver Gruppe und Zielnukleobase effizient ablaufen kann.

In Kapitel 3 wird die punktspezifische Modifizierung der C4-Position von Cytosin beschrieben, welche unter Hydrogensulfit-Katalyse abläuft. Durch Variation der Länge des Linkerbausteins und den Einbau von Hydrazinderivaten in die reaktive Gruppe konnte die Verbrückungseffizienz beider Stränge auf bis zu 60% gesteigert werden. In den Kapiteln 4 und 5 wurden ein Bromacetatderivat sowie eine Acrylamidgruppe dazu verwendet, Guanin und Thymin punktspezifisch zu alkylieren. Zur Reaktionskontrolle wurden die Reaktionsprodukte enzymatisch verdaut, was die einzelnen Nukleotide liefert bzw. in Falle einer erfolgreichen Verbrückung auch das entsprechende Dinukleotid. Die ursprüngliche Reaktionsmischung sowohl die Abbauprodukte wurden in Folge mit MALDI-TOF MS analysiert.

Die Verwendung eines spaltbaren Linkerbausteins ermöglicht die Trennung der kovalent verbrückten Stränge unter geeigneten Bedingungen. Die dabei an der modifizierten Nucleobase verbleibenden funktionellen Gruppen können für weitere Modifizierungen genutzt werden. In Kapitel 3 wird die oxidative Spaltung einer Diolgruppe im Linkermolekül zu einer Aldehydgruppe gezeigt, welche anschliessend durch reduktive Aminierung an ein Aminofluoreszinderivat gebunden wird. In den Kapiteln 4 und 5 werden Linkerbausteine verwendet zur Verbrückung verwendet, welche über spaltbare Disulfidbrücken verfügen.

Schliesslich wird in Kapitel 6 eine allgemeine Methode für die punktspezifische Funktionalisierung von Nukleinsäuren für fluoreszenzspektroskopische Anwendungen vorgestellt. So konnte eine Thiolgruppe punktspezifisch an eine von 14 Guaninbasen in einer 50 Nucleobasen umfassenden Zielsequenz eingefügt werden. Im Anschluss wurde diese Thiolgruppe mit einem Maleimidosulfo-Cy3 Farbstoff gekoppelt und das Produkt mittels MALDI-TOF, RP-HPLC and PAGE charakterisiert. In einem abschliessenden Experiment wurde der Cy3-modifizierte Strang mit einem so genannten Cy5-Akzeptorstrang hybridisiert. Daraufhin konnte eine Zunahme der Fluoreszenzemission des Farbstoffes auf dem Akzeptorstrang beobachtet werden, was mit grosser Wahrscheinlichkeit auf einen Energietransfer zwischen den beiden Farbstoffen zurückzuführen ist.

List of content

List of abbreviations

A	Adenosine
[γ-^{32}P]-ATP	P^{32}-adenosine triphosphate
ANP	3-amino-3-(2-nitrophenyl)propionic acid
NBS	N-bromosuccinimide
Boc	Tert-butyl-carbonate
CTAB	Cetyl trimethylammonium bromide
C	Cytidine
DNA	Deoxyribonucleic acid
DCM	Dichloromethane
DCC	N,N'-Dicyclohexylcarbodiimide
DIPEA	N,N-Diisopropylethylamine
DMTMM	4-(4,6-Dimethoxy[1.3.5]triazin-2-yl)-4-methylmorpholinium chloride hydrate
HATU	N-[(Dimethylamino)-1H-1,2,3-triazolo-[4,5-b]pyridin-1-ylmethylene]-N-methylmethanaminium hexafluorophosphate N-oxide
EDC	1-(3-Dimethylaminopropyl)-3-ethylcarbodiimide hydrochloride
DMF	N,N-Dimethylformamide
DMT	Dimethoxytrityl
DMAP	4-Dimethylaminopyridine
DTT	(2S,3S)-1,4-Disulfanylbutane-2,3-diol
PDS	2,2'-Dithiodipyridine
ESI MS	ElectroSpray Ionization Mass Spectrometry
EDC	N-Ethyl-N′-(3-dimethylaminopropyl)carbodiimide hydrochloride
EDTA	Ethylenediaminetetraacetic acid
FRET	Förster resonance energy transfer
G	Guanosine
NHS	N-hydroxysuccinimide
ICL	Interstrand Cross-Link
MALDI_TOF	Matrix-Assisted Laser Desorption/ionization ─ Time Of Flight
MeOH	Methanol
mBBr	Monobromobimane
PNA	Peptide Nucleic Acid
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction

RP-HPLC	Reverse Phase High-Performance Liquid Chromatography
RNA	Ribonucleic acid
SRS	Short Recognition Sequence
T	Thymine
TAL	Transcription-Activator Like effector
TEAA	Triethylamine acetate
TFA	Trifluoroacetic acid
PPh_3	Triphenylphosphine
TFO	Triplex Formation System
TCEP	Tris(2-carboxyethyl)phosphine

CHAPTER 1

1. Introduction

1.1 DNA structure and function

Ribonucleic acids, i.e. DNA and RNA, are linear polymers that are composed of four monomeric units in different order. Each monomeric unit, the nucleotide, consists of nucleobase, a sugar ring and a phosphate group, and is connected to the next nucleotide by a phosphodiester bond. The primary sequences in ribonucleic acids are defined by adenine, guanine, cytosine and thymine (in case of DNA) or uracil (RNA) nucleobases. Due to the electronic structure of the nucleobases, the single stranded polymers can form strong hydrogen bonds that result in formation of secondary structures. Under physiological conditions two DNA single strands wind around a common axis and form a stable helical structure. The vast majority of naturally occurring DNA exists in the canonical B-form duplex (**Fig. 1.1**). The duplex is roughly 20 Å in diameter with the two strands running antiparallel to each other. The hydrophobic nucleobases form the core of the helix and stabilize the structure by hydrogen bonds and π-stacking interactions between the aromatic rings. In 1953, Watson and Crick interpreted and described the possible combination of hydrogen bonds in the helix[1]. Accordingly, the hydrogen bonds formed between C – G and A – T are denoted as "Watson-Crick hydrogen bonds" (**Fig. 1.2**). The base planes are almost perpendicular to the helical axis and placed 3.2 Å apart from each other. There are ten bases per pitch in B-form DNA. The sugar-phosphate backbone that is pointing to the periphery of the helix, decreases the electrostatic repulsion between negatively charged phosphate groups of opposite strands and at the same time increases the solubility of the polymer by hydration with water molecules.

Figure 1.1. Structure of A-, B-, and Z- form of DNA. The picture was taken from (http://commons.wikimedia.org/wiki/File:A-B-Z-DNA_Side_View_Transparent.png). Copyright was granted under the terms of the "GNU Free Documentation License" (http://commons.wikimedia.org/wiki/GNU_Free_Documentation_License).

The DNA helix can also adopt the A-form. This is a right-handed double helix with antiparallel chains. However, the helical axis in the A-form lies on the periphery of the macromolecule in contrast to B-form, where the axis passes through the center of the structure. The planes of the nucleobases are not perpendicular to the central axis, but they form an angle of 20°, and the bases of the basepairs are not coplanar to each other. To reach this conformation, the C3'-atom of the sugar moiety changed its conformation from the *exo*- (B-form) to the *endo*-configuration (A-form). All of these properties result in tighter packing of mononucleotides in the macromolecule and the distance between bases is reduced to 2.4 Å.

Precise X-ray analysis[2] confirmed the existence of a left-handed helix which became known as Z-DNA**.** The repeating unit in the Z-form of DNA consists of two base pairs, in contrast to the B- and A-forms; as a result, the line connecting the phosphate groups takes a sharp turn and assumes a zigzag shape. In contrast to B-DNA, the helix diameter is with 18 Å much smaller.

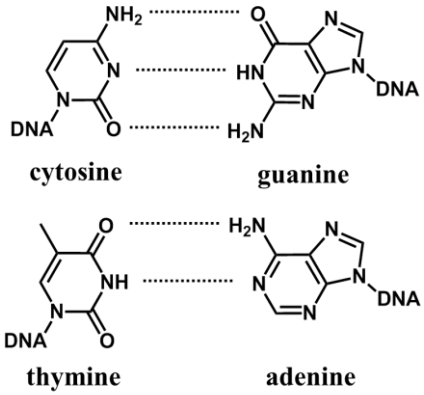


Figure 1.2. Watson - Crick model of formation of hydrogen bonds in the DNA helix.

In addition to double helix conformation, DNA can form triplex[3], i-motif[4], hairpin[5] and G-quadruplex[6] structures forming additional non-canonical hydrogen bonds. Three DNA single strands can bind in a parallel (C and T rich region) or antiparallel orientation (G- and T-rich regions) *via* what is commonly known as Hoogsteen hydrogen bonds (**Fig. 1.3**).

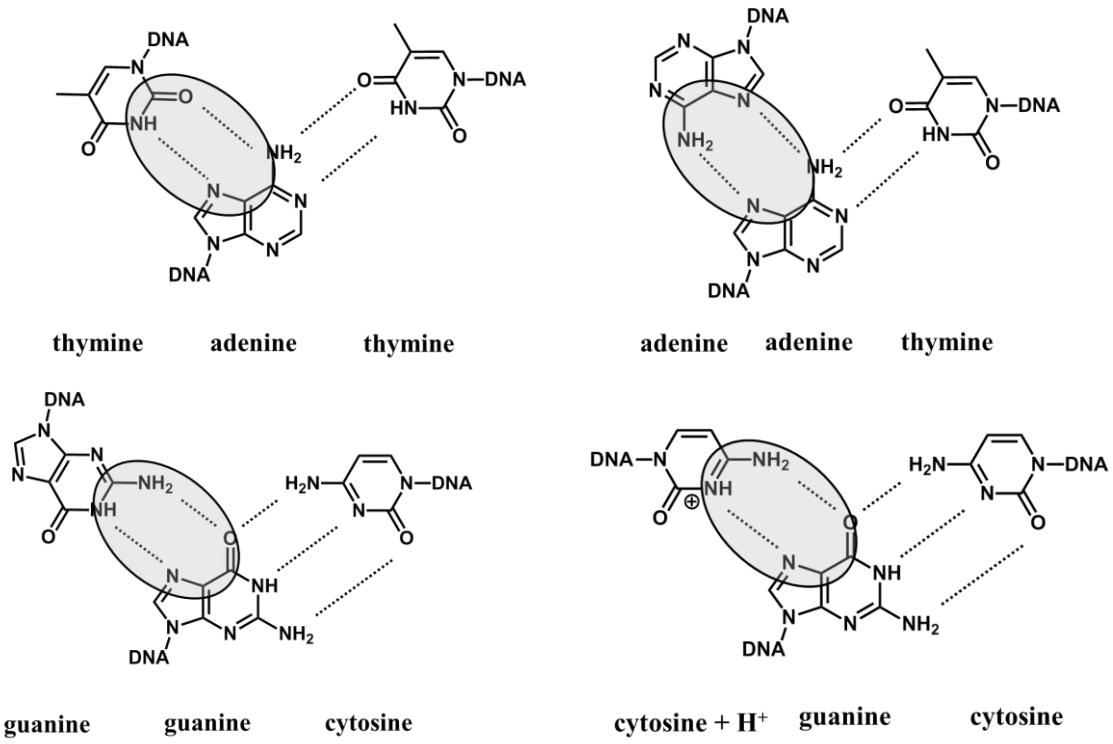
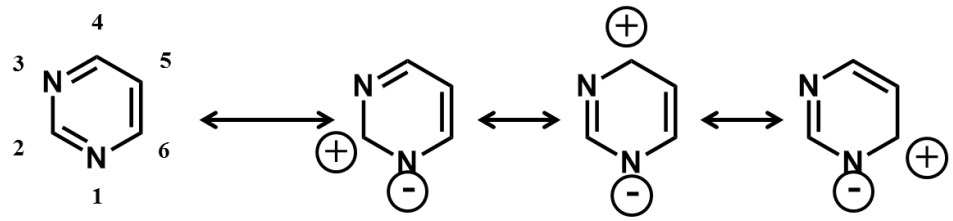


Figure 1.3. Hoogsteen – type hydrogen bonds (in circles) T-AT, A-AT, G-GC and C(H^+)-GC triplets within a triple-helix motif. The picture was adapted from [3].

1.2 Chemical properties of nucleotides and possible site for modification.

Chemical properties of oligonucleotides highly depend on the chemical reactivity of the single nucleosides. Each nucleotide is characterized by a specific electronic structure that includes carbon, nitrogen and oxygen atoms. Nevertheless, purine and pyrimidine nucleobases have similar features. All ring nitrogens in the nucleoside structure can be divided into two groups – pyridine and pyrrole like nitrogens. The first type of reactive sites is represented by pyrrole nitrogens that form three σ-bonds between two carbons and one hydrogen, hence the orbitals lie in the plane of the aromatic ring. The electron lone pair of nitrogen is involved in formation of the π-conjugated system with the neighboring atoms and its orbitals lie perpendicular to the plane of the heterocycle[7]. The electronic structure of the exocyclic amino group nitrogens is similar. Thus, this type of nitrogen atoms will be affected by electrophilic reagents when the attack is directed at a right angle to the plane of the heterocyclic base due to orientation of the orbital of the lone-pair electrons of nitrogen[8]. In contrast, the pyridine nitrogens form only two σ-bonds with carbons and the potentially reactive unbound pair of electrons also lies in the plane of the heterocyclic ring. Accordingly, the electrophilic attack must be directed toward this plane.

Both types of nitrogens participate in mesomeric structures and have an influence on the distribution of electron density in the π-conjugated system. In pyrimidine nucleotides this distribution results in electron deficiency of the carbons at positions 2, 4, and 6 (**Scheme 1.1**). These carbons are vulnerable in reactions with nucleophilic reagent. Only the carbon atom at position 5 has electrophilic properties.



Scheme 1.1. Resonance structures of an unsubstituted pyrimidine ring. Picture was adapted from [9].

The distribution of electron density in the purine rings and accessibility of carbons to take part in chemical modifications are more challenging to evaluate. According to resonance structures, only the C8 and C6 positions are accessible for modification. The other carbon atoms are relatively inert in chemical reactions and can only be modified under more harsh conditions.

In the above discussion we assumed that the nucleobases are present in their neutral form. Under reaction conditions and hence at different pH values each of the bases can donate or accept protons due to their acid-base properties resulting in charged forms with different reactivities. For example, transamination of cytosine is initiated by protonation of the N3 position at slightly acidic conditions.

1.3 The two main determinants of DNA helix modification: Steric factors and polyfunctionality

The chemical properties of DNA are determined by the chemical structure of the mononucleotides such as N-glycosidic bonds, carbohydrate moieties, terminal phosphate and heterocyclic rings (**Fig. 1.4**). At the same time two factors govern the strategy of DNA modification[10].

The first factor refers to the accessibility of the reactive group at the nucleoside. Indeed, reaction between nucleoside and incoming reagent requires sterically correct positioning of reagent and nucleotide relative to each other. The DNA helix is a rather rigid structure where heterocyclic planes are stabilized by hydrogen bonding and stacking interaction. This means that in appropriate reaction, the nucleophilic or electrophilic reagent should be placed in plane with

the heterocyclic ring or in a right angle to it. Hence only part of the contacts between reagent and the nucleotide will result in a reaction. In addition to Watson-Crick hydrogen bond formation[11], the DNA helix structure is stabilized by π-stacking interaction of the aromatic systems of the nucleobases. This type of interaction hinders a reagent's attack if it is directed at a right angle to the plane of the base. It will lower the speed of nucleoside modification in DNA compared to modification of mononucleotides. Consequently, in order to ensure sufficiently fast reaction rates, nucleic acids are modified using reagents that are added in excess. The steric factor plays an important role in site-specific modification that is initiated by binding of a third DNA strand to the duplex region. This places the target nucleobases in the center of the DNA, where they are additionally protected by Hoogsteen bonds and π-stacking. This steric hindrance makes them inert for most types of modifications. Only reactions on the side of the groove or by targeting the sugar moiety[12,13] are efficient in this strategy of site-specific modifications.

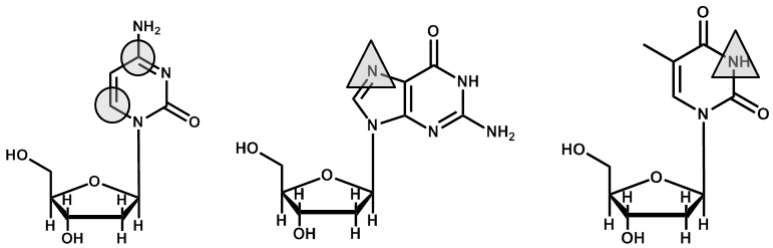


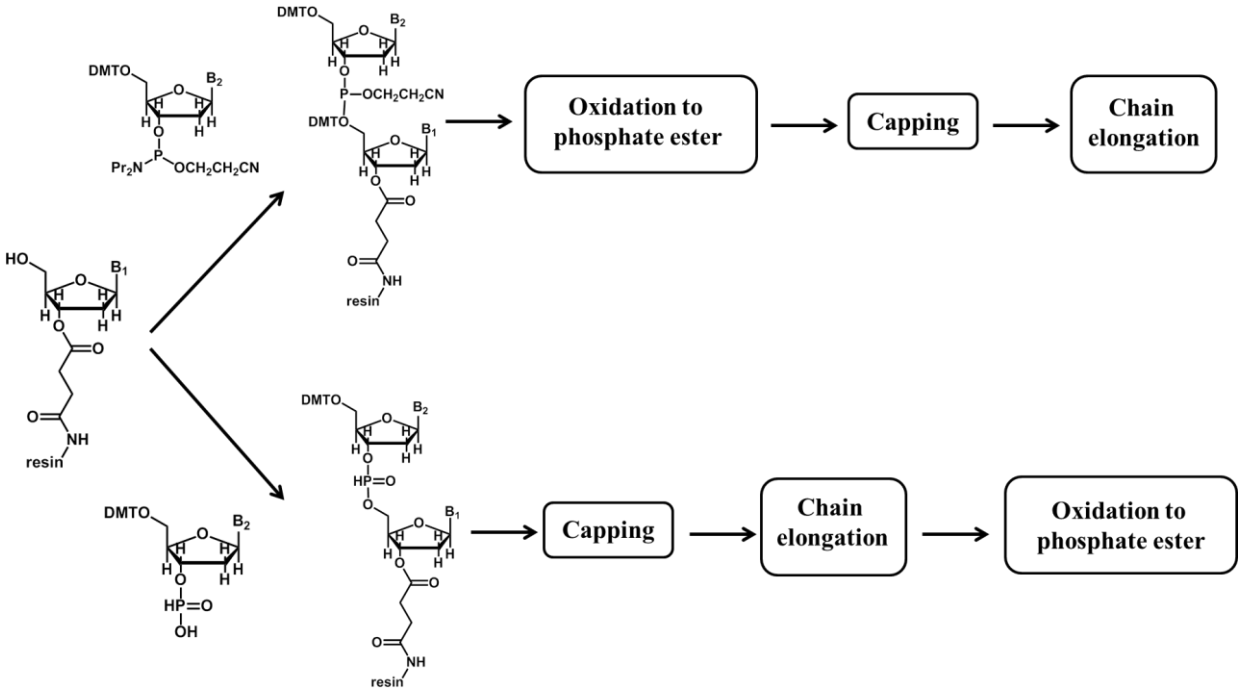
Figure 1.4. Reactive sites in heterocyclic rings of nucleobases that were used for modifications. Triangles highlight nitrogen atoms that are preferential positions for electrophilic attack. Circles denote carbon atoms that are affected by nucleophilic agents.

Another factor that determines the reactivity of DNA is the polyfunctionality of the polymer structure. The presence of a large number of identical and structurally similar reactive sites in a nucleic acid requires careful design of the reactive chemical. The reactive group of the reactive sequence should allow the modification of one nucleobase in the presence of the three others. At the same time the huge amount of identical nucleotides in different positions of the target sequence requires a new strategy for the localization of the reactive group close to the target nucleotide.

1.4 Methods of DNA modifications: Introducing an unnatural nucleoside during solid-phase synthesis.

The most common strategy for the synthesis of modified oligonucleotides is up to now automated solid-phase synthesis[14]. The general idea of this strategy relies on the immobilization of protected mononucleotides on a surface through a cleavable linker and stepwise synthesis of the growing oligonucleotide strand[15]. Two methods are used for the attachment of the next base

to the immobilized sequence, i.e. either *via* phosphoramidite or *via* H-phosphonate reactive groups[16] (**Scheme 1.2**). With both strategies, the chain is elongated by coupling protected mononucleotide building blocks to the 5'-hydroxy group of the immobilized strand with formation of phosphite-triester (in phosphoramidite method) or phosphortriester intermediates (in H-phosphonate method). The major difference of the two strategies is hidden in the step of oxidation of this intermediate. In the phosphoramidite strategy the phosphite-triester is oxidised to a phosphate ester after each step of conjugation, while in the H-phosphonate method the oxidation to phosphate esters happens after completion of the chain assembly in a single step. The last step is similar for both pathways and includes full deprotection of the oligonucleotide sequences and their cleavage from the solid support.



Scheme 1.2. Solid-phase synthesis of an oligonucleotide primer using phosphoramidite and the H-phosphonate strategy. "DMT" is a dimethoxytrityl protecting group. Picture was adapted from [16].

A great variety of modified mononucleotide building blocks are commercially available. Building blocks that carry a primary aminofunction, alkyne or chloride precursor group are among the most popular. Additional advantages of this method are the high yields after each coupling step, low prices and reproducible protocols. Nevertheless, even a coupling efficiency of 98-99% per step leads to significant decreases in yield with increasing chain length. Accordingly, the method is only economic and feasible for shorter oligonucleotides and prompts researchers to apply alternative approaches for the synthesis of long nucleic acids.

1.5 Methods of DNA modifications: Enzymatic techniques.

An alternative for automated solid-phase synthesis are modified mononucleotide building blocks that can be inserted in the oligonucleotide strand using different enzymatic techniques[17-19]. The most popular protocols utilize DNA polymerase I[20] or terminal deoxynucleotide transferase[21]. The first enzyme plays a key role in polymerase chain reaction (PCR) technology[22]. According to this protocol, the modified oligonucleotide is prepared from two strands (**Fig. 1.5**). The first strand consists of an unmodified sequence, while the second strand contains the modified mononucleotide in a specific position and is usually prepared by solid-phase synthesis. In the next step, the unmodified sequence is annealed to the modified template. The modified template is used in the PCR as a primer and Taq polymerase elongates this primer to give the final long modified strand. When the terminal transferase is used no template is needed as the enzyme directly adds modified nucleoside to the 3'-end of the oligonucleotide[23].

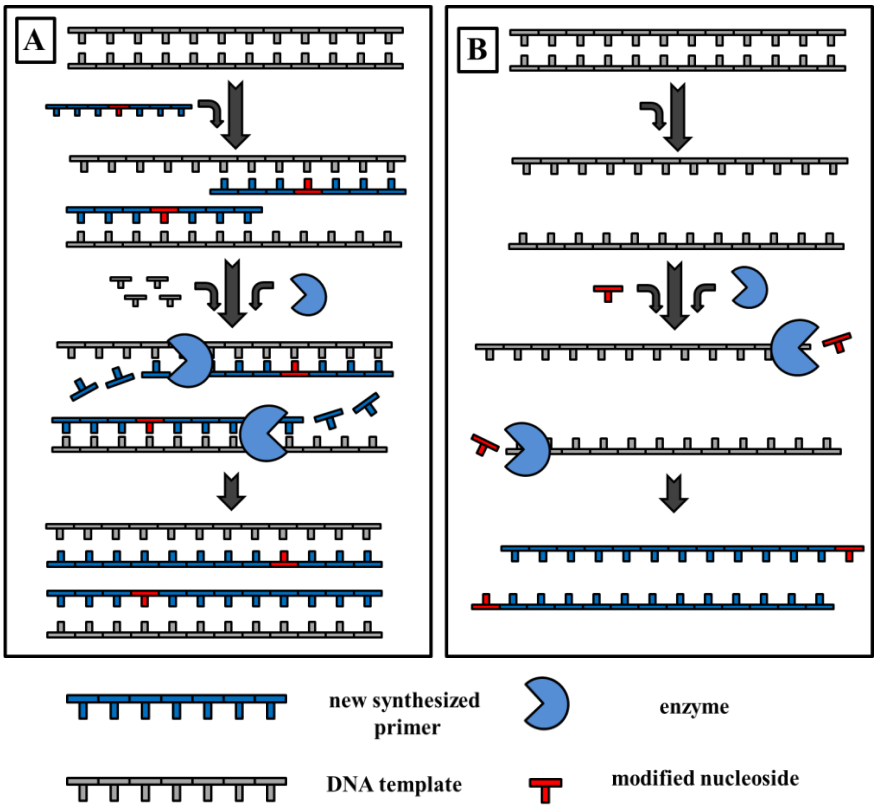


Figure 1.5. General scheme of PCR technology. **A:** modified nucleotide is introduced in the short primer during solid phase synthesis. After annealing to complementary template, full length of modified sequence is produced by the addition of Taq polymerase. **B:** modified triphosphate nucleoside is added to solution and enzymatically attached to the growing oligonucleotide chain by terminal deoxynucleotide transferase.

However, the range of nucleosides (XTPs) that can be successfully incorporated into an oligonucleotide by PCR is severely restricted due to limitations in size and configuration of the polymerase active center[24]. Accordingly, for purine nucleotides, only the N6, N7 (as 7-deaza-derivatives) and C8 positions are suitable for incorporating side chains into nucleic acid by PCR (**Fig. 1.6**). The pyrimidine nucleosides are even more limited in the position of an additional side chain and only the C5 position can be modified with a long spacer arm.

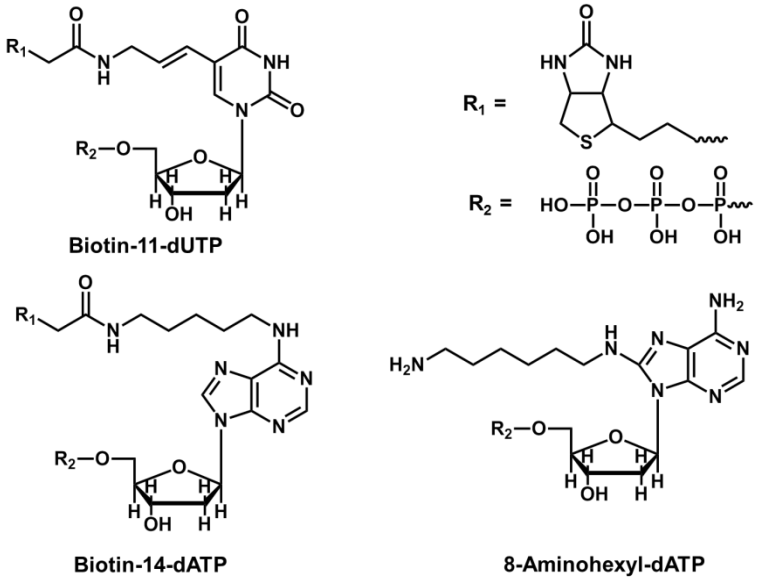


Figure 1.6. Examples of nucleoside triphosphate derivatives that can be incorporated into oligonucleotides by PCR. Picture was adapted from[25].

1.6 Methods of DNA modifications: Unspecific modifications with small organic chemicals.

Non-specific post-synthetical modification provides an alternative strategy for the labeling of oligonucleotides in comparison with automated solid-phase synthesis. This approach has no limitation regarding the length of the oligonucleotides and is based on the individual chemical reactivity of each type of nucleoside. This approach requires the design of reactive molecules that specifically react with only one type of nucleosides in the presence of the three others. The disadvantage of this method is that all nucleosides of one type within a given sequence can be modified, regardless of their position (**Fig.1.7 A**).

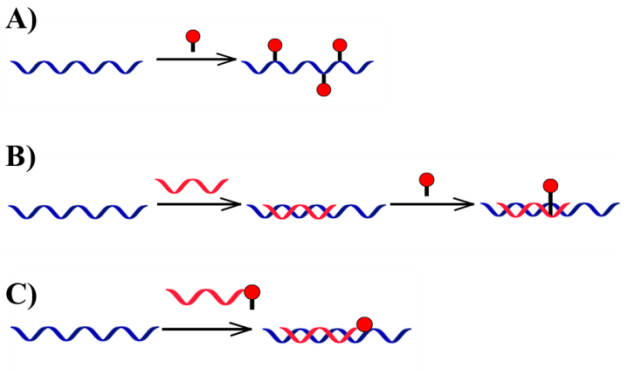


Figure 1.7. Three strategies for post-synthetic labeling of oligonucleotides. **A:** Unspecific modifications with small organic compounds. **B:** None-specific interstrand cross-link formation. **C:** Site-specific covalent cross-link formation in DNA duplexes.

In the last 50-years a lot of efforts were focused on non-specific covalent labeling of oligonucleotide sequences that results in different synthetic routes (**Table 1.1**). One of the most popular chemical modifications is based on transamination of cytosine at position N4 catalyzed by sodium hydrogensulfite[26] (**Table 1.1, entry 3**). In this reaction a sulfonate group is added to the 5,6-double bond of cytosine in single-stranded DNA resulting in a 6-sulfo-cytosine derivative. This activates the C4 atom of cytosine and makes it a good target for nucleophilic replacement. Under physiological conditions, even weak nucleophiles such as water can replace the amino group leading to formation of uracil[27]. Shapiro and Weisgras[28] demonstrated that in presence of nucleophilic compounds such as amine- or hydrazine-derivatives, 6-sulfo-cytosine intermediates are efficiently transformed into derivatives with a new side chain at the C4 position. Hence this reaction is a straightforward method to incorporate amino-modified reactive groups such as thiols or alkynes into nucleic acids through cytosine bases[29].

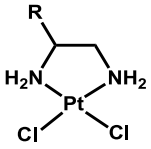
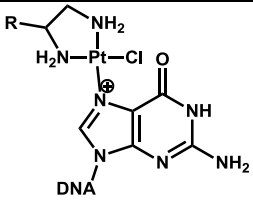
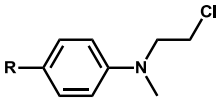
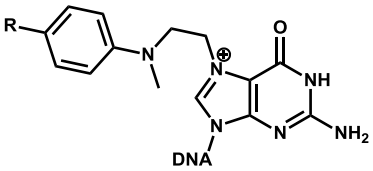
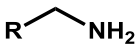
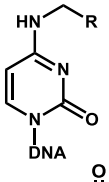
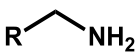
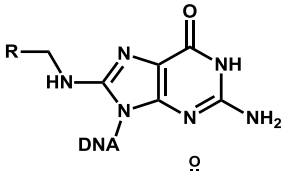
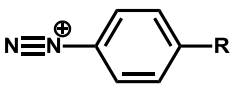
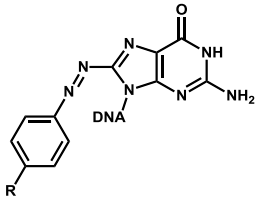
A more general method for the labeling of oligonucleotides that is suitable for both purine and pyrimidine rings was demonstrated by Okamoto and Saito[30]. They activated nucleobases of DNA and RNA with bromine to produce reactive intermediates capable of coupling with nucleophiles. Bromination occurs at the C8 position of guanine residues and at C5 of cytosine. In RNA the C5 position of uracil is a less preferable target for modification due to the electron-withdrawing effect of the C4=O carbonyl group. Adenine is less reactive in both types of biomolecules due to its high ionization potential and large energy barrier for initiating of reaction. The reactive intermediate is reacted *in situ* with nucleophiles and forms amino-labeled bases[31,32]. Either an aqueous solution of bromine or the compound N-bromosuccinimide can be used for this reaction (**Table 1.1, entry 4**). Opposite to transamination of cytosine, the sites of

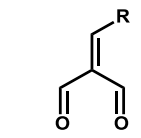
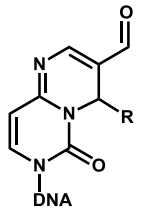
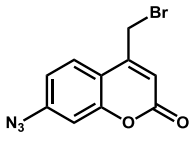
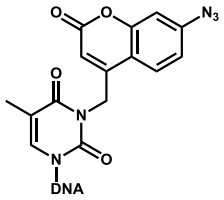
derivatization using bromine activation are not involved in hydrogen bonding during base pairing, hence the hybridization ability of the modified nucleobase will be maintained.

Rothenberg and Wilchek[33] proposed to use a diazonium-containing molecule for the specific modification of the guanine C8 position (**Table 1.1, entry 5**). They treated a single stranded M13 DNA mp19 plasmid with p-diazobenzoyl-biocytin that was synthetized independently. The biotin-labeled plasmid was detected in dot hybridization experiments using streptavidin and alkaline phosphatase. Unfortunatelly, they did not published any following step with optimization protocol or efficiency of hybridization.

However, as mentioned above, all of these post-synthetical methods have the major drawback of being not site-specific leading to modification of numerous nucleobases of the same kind in the sequence.

Table 1.1. Small organic reactive agents or metal-complexes for non-specific DNA modification.

Entry	Reagent	Reagent 2	Residue	Product	Ref.
1			Guanine		34
2			Guanine		35
3		$NaHSO_3$	Cytosine		36
4		NBS	Guanine		30
5			Guanine		33

6		Cytosine		37
7		Thymine		38

1.7 Non-specific interstrand cross-link modifications between complementary DNA strands

The first investigations of site-specific modification of oligonucleotides were focused on the formation of covalent cross-links between complementary DNA strands[39,40] (**Fig. 1.7 B**). In the case of single stranded DNA or when using mismatched sequences, the cross-link reagent was unreactive or led to side reactions (**Table 1.1**).

The first cross-linking reagents used for modification and conjugation of macromolecules consisted of two similar reactive groups and a spacer between them[41]. Cross-linking reagent can also contain additional groups, i.e. for intercalation into the DNA helical structure or for fluorescent imaging of the co-labeled structure. The N,N-(2-chloroethyl) amines represent the first class of interstrand cross-linking agents[42] (**Scheme 1.3, A**). Because the two identical 2-chloroethyl reactive groups they are usually denoted "nitrogen mustards". The site-specificity of this class is given by these 2-chloroethyl groups that preferentially alkylate the N7 position of guanine and direct modifications to the -GpC- rich regions in the DNA helix. Modification of the other sites in purine nucleobase, e.g. O6, O4, N3 etc., can be observed. Nitrogen mustards are widely used in the chemotherapy of chronic lympholeucosis and other type of cancer[43].

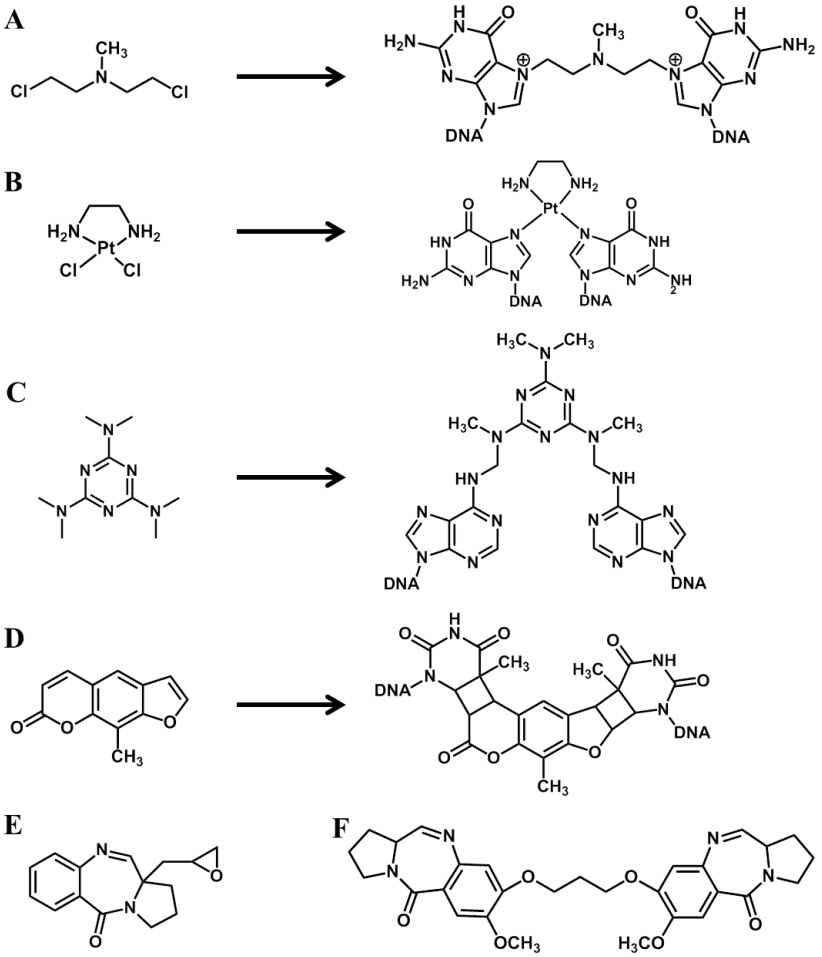
The second large class of bifunctional reagents is represented by dichloroplatinum(II)-based metal complexes[44] (**Scheme 1.3, B**). The reactive site of these chemicals contains a Pt-center in square-planar coordination environment. Evidence suggests that the two chloro ligands of the dichloroplatinum(II)-complexes are exchanged via a two-step hydration process[45,46]. Replacement of the chloro ligands by aqua ligands adds two positive charges to the formerly neutral Pt(II)-complex, which can interact with negatively charged oligonucleotide sequences via electrostatic interaction followed by coordination to a nucleotide. In contrast to nitrogen mustards, this class of reagents is rather specific (> 90% of DNA platination sites[47]) for purine-N7 modification[48]. Mononuclear Pt-complexes can only modify two neighboring guanines. By

increasing the number of platinum centers in the molecule and by designing longer spacers between them cross-link formation can be also achieved between guanines further apart and hence the reagent can be tuned to target specific sequences[49,50].

The pyrrolo[2,1-*c*][1,4]benzodiazepines[51] (PBDs) represent a small class of cross-linking agents The core of these molecules contains a planar ring with a large π-conjugated system usually consisting of 11 conjugated atoms[52] that allows to form hydrogen bond between C11 hydroxyl-group (**Scheme 1.3, E**) from benzodiazepine and exocyclic N2 amine of deoxyguanosine in the minor groove of DNA duplex followed by covalent cross-linking of two strands[53-56]. Hartley et al.[57,58] synthesized a benzodiazepine dimer and demonstrated a 50-fold increased cross-linking efficiency compared to dichloroplatinum(II)-complexes (**Scheme 1.3, F**).

In spite of the high chemical reactivity of compounds with chloro-ligands as leaving groups and their efficiency in cross-link formation in -GpC- rich sequences, molecular biology applications and clinical trials suffer from difficulties in specific drug delivery to the target macromolecule. To increase drug delivery of the cell, several trials have been made to deliver an inactive precursor to the cell followed by activation and immediate binding to the target site in the nucleic acid. Hexamethylmelamine-derivatives are the first example for the successful application of this strategy[59] (**Scheme 1.3, C**). Little is known about mechanism of mechanism of it activation and its relationship to DNA alkylation. Jackson et al.[60] proposed formation of reactive imminium intermediate that is responsible for DNA alkylation.

The strategy of photoactivation of a precursor to yield the reactive compound provides a more precise control for *in vivo* cross-link formation that does not depend on the physiological surrounding. Psoralen derivatives represent the most popular class of photoactivated compounds[61] (**Scheme 1.3, D**). Their aromatic core contains a furan ring conjugated to a coumarin residue. Cross-linking psoralen to a DNA duplex requires three steps. In the first step, due to the unique structure, psoralen preferentially intercalates into the 5'-AT-3' motif of the duplex. Upon irradiation, the 4',5'-furan double bond or the 3,4-pyrone double bond form a cyclobutane ring with the 5,6-double bond from one of the thymine base[62]. Absorption of a second photon results in a second cyclobutane cyclization with a neighboring pyrimidine affording interstrand cross-link formation.



Scheme 1.3. Small organic reactive agents for non-specific covalent interstrand cross-link formation in oligonucleotide sequences.

1.8 Site-specific covalent cross-link formation in DNA duplexes.

The preferable affinity of benzodiazepine and psoralen molecules to specific oligonucleotide motives that is based on attachment of the molecules within the helix is however not sufficient to reach real site-specificity. Hence to achieve this specificity the reactive molecule was attached to a guiding moiety, e.g., an oligonucleotide strand or minor groove binding regents[63-65], etc (**Fig. 1.7 C**).

Incorporation of DNA single strands complementary to the desired position in the target template has been reported in systems with different reactive groups[66,67,35] (**Table 1.2**). The N7 nitrogen atom of guanine is a preferable target for modification due to its high nucleophilicity and position in the major grove of DNA[68]. In contrast, the N7 nitrogen atom from adenine is less reactive towards this type of modification. Additionally, N7-alkylated guanine is efficiently depurinated from DNA in presence of piperidine leading to cleavage of the long DNA molecule

into two parts[69]. Dervan at al. attached a 2-bromoacetate reactive group to an oligonucleotide guide sequence that should form a triple helix with the G-rich region of a long DNA molecule[70]. After annealing the reactive group should be localized close in space to the target guanine and be able to modify it. Indeed, it was demonstrated that after piperidine cleavage only two new DNA pieces were formed[71] strongly suggesting the modification of just a single guanine in the sequence. In the control experiment, where a G-rich DNA helix was unspecifically modified with the same reactive group, a great number of new DNA pieces were formed. Another example for the controlled site-specific cleavage of a long DNA strand makes use of a chemical nuclease. Attachment of a DNA guiding sequence to an Fe(II)-EDTA complex increased the cleavage specificity of this complex towards the DNA phosphate backbone[72]. Similar results were demonstrated with other types of chemical nucleases such as Fe(II)-phthalocyanide[73], lanthanide(III)-iminotripyridines[74], Eu(III)-porphyrin[75], Zn(II)-imidazole[76], and Cu(II)-chelate[77,78] complexes.

In the last years the focus of cross-link formation reactions was shifted towards the formation of stable final products. Luedtke et al. introduced an artificial nucleoside that contains a 2-chloroethyl reactive group to the middle of a guide sequence[79]. After annealing they demonstrated formation of an interstrand cross-link with the cytosine opposite the reactive group with a yield of almost 50%.

In contrast to light-activated compounds, Madder et al.[80,81] focused on formation in situ highly reactive compounds. In the initial experiments the precursor of a reactive group that contained a furan-moiety attached to a guiding sequence was transformed to the reactive enal by oxidation with NBS. After annealing and activation, fast and efficient alkylation of the opposite cytosine has been achieved.

Many efforts were undertaken to site-specifically label target templates with Pt-drugs due to their high importance in anticancer medicine. One of the possible ways to reach this requires incorporation of a cisplatin derivative to the recognition sequence during solid-phase synthesis[82]. The reactive molecule was annealed to the complementary strand that contains guanine at the appropriate position and ICL duplex was formed with efficiency of 58%[83].

Photochemical reagents provide the most precise control of cross-link formation. They are inert under physiological conditions, and cross-linking is induced only upon irradiation with light. Due to this fact, this method does not require a chemical for activation and can be applied under physiological conditions. To test this assumption, Vasquez[84] attached a photochemically reactive psoralen derivative (4'-(hydroxymethyl)-4,5',8-trimethylpsoralen) to the oligonucleotide sequence that binds with high affinity to the G-rich region in the hamster adenine phosphoribosyltransferase (APRT) gene. After annealing and irradiation, they detected high

efficiency (> 60%) of ICL formation between oligonucleotide with psoralen moiety and a target site of the APRT gene.

Kjems et al.[85] demonstrated that site-specific labeling of long oligonucleotides does not require a complementary donor strand but can be also achieved with alternative structures, i.e. a four-way junction (**Fig. 1.8**). As a target they used an RNA single strand and tested the possibility of its modification through the 2'-OH group of the sugar moiety. Two additional unmodified DNA strands were complementary with one of their ends to the RNA sequence at both sides of the position of the target nucleoside. The other ends of the additional DNA strands were complementary to the donor DNA strand creating a four-way junction as depicted in **Fig. 1.8**. A carboxylic group was coupled to the 2'-OH group of the target RNA strand in the presence of 4-(4,6-Dimethoxy[1.3.5]triazin-2-yl)-4-methylmorpholinium chloride. Unfortunately, the efficiency of cross-link formation was low and maximum yield was around 5%.

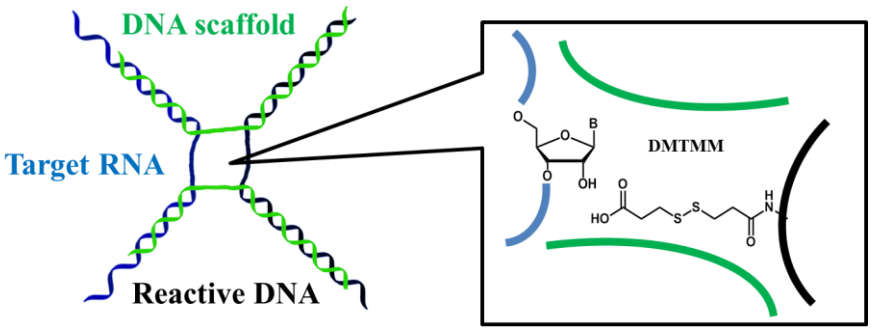
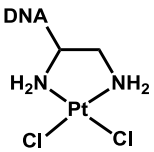
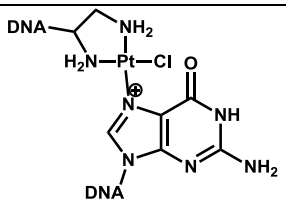
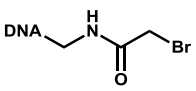
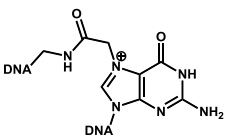
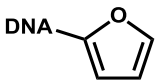
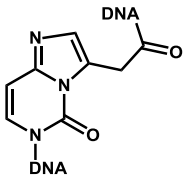
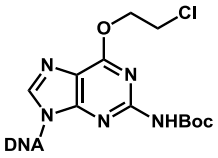
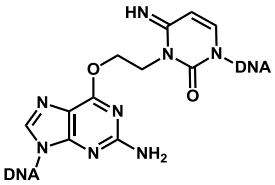
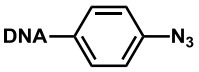
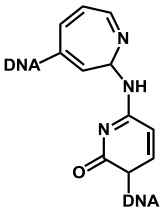
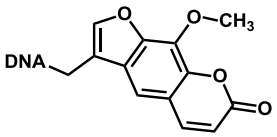
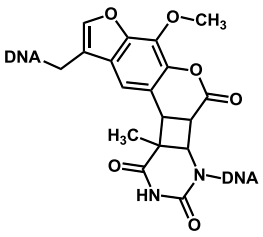
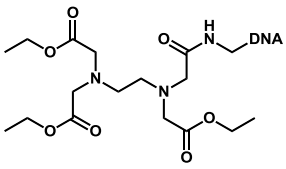


Figure 1.8. Four way junction labeling strategy for modification of the 2-OH'-group in an RNA template. Blue line represented target RNA. Green lines represented additional DNA scaffolds. Black line represented reactive sequence. Picture was adopted from[85].

Table 1.2. Reactive groups that were attached to guide sequences for site-specific modification of DNA oligonucleotides.

Entry	Reagent	Reagent 2	Residue	Product	Ref.
1			Guanine		83
2			Guanine		71

3		NBS	Cytosine		80
4		λ = 365 nm	Cytosine		79
5		λ = 360 nm	Cytosine, Thymine		86
6		λ = 365 nm	Thymine		84
7		Fe^{2+}	Phospho diester bond	Hydrolysis of phosphodiester bonds	72

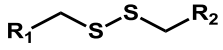
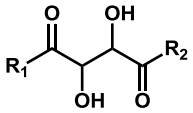
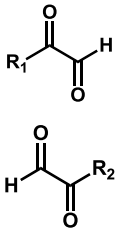
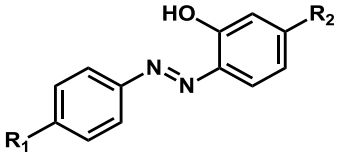
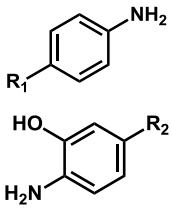
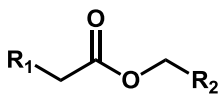
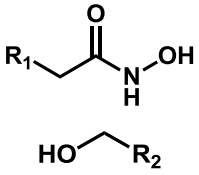
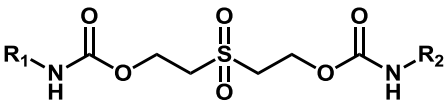
1.9 Chemoselective cleavable systems.

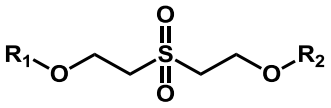
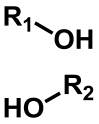
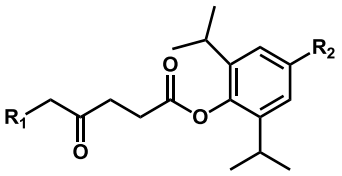
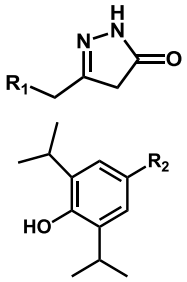
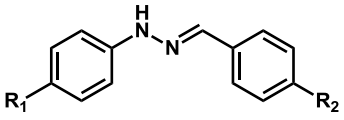
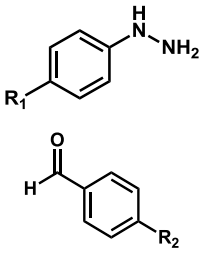
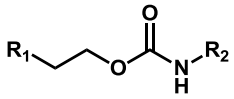
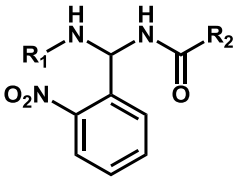
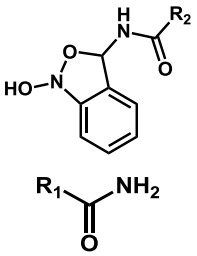
The introduction of a specific linker between reactive sequence and the target oligonucleotide, which can be cleaved chemoselectively while maintaining the stability of the oligonucleotides, is an attractive strategy to site-specific DNA modification (**Table 1.3**). The cleavable part of the linker should be stable under the conditions of cross-link formation (reactive groups, buffers, reducing agents). At the same time, it must be susceptible to chemoselective cleavage conditions.

The ability to cleave a cross-link bridge also opens up the possibility to transfer a functional group from the recognition sequence to the target oligonucleotide[87]. Examples of structures that have this capability include the disulfide and glycol moiety. Disulfide bonds are stable under different conditions but can be easy reduced to thiol groups by e.g. TCEP or DTT via disulfide exchange processes[88,89] (**Table 1.3, disulfide bridge**). The resulting thiol groups can be used for a great variety of chemical modifications such as the attachment of maleimide[90-92], iodoacetate[93] or sulfone moieties[94,95].

Tartaric acid derivatives[96] or similar compounds containing glycol groups can be easily cleaved by oxidation with sodium periodate at slightly acidic pH leading to give two aldehydes[97](**Table 1.3. glycol bridge**). It is important to note that cleavage of both types of cleavable linkers will add the same functional group, i.e. a thiol or an aldehyde to both the reactive sequence and the target strand[98]. It should be kept in mind that under these conditions the 3'-sugar moiety of RNA is also susceptible to oxidative cleavage forming additional aldehyde groups[99]. As the third example, covalent bridges containing diazo-bonds within their structures can be specifically cleaved with dithionite[100] (**Table 1.3. Bisaryl-diazo bridge**). Sodium dithionite reduces the diazo linkage and generates a primary amino-group on both fragments of the system. The reaction is usually carried out under alkaline conditions.

Table 1.3. Cleavable linker for interstrand cross-linking for use in biomolecular applications.

Name	Linker	Cleavage conditions	Products	Ref.
Disulfide bridge		DTT, TCEP		89
Glycol bridge		$NaIO_4$		96
Bisaryl-diazo-bridge		$Na_2S_2O_4$		100
Ester bridge		NH_2OH		101
Sulfone bridge with amino-group		pH = 9.6		102

Sulfone bridge with hydroxy-group		pH = 11.6		103
Levulinoyl bridge		$N_2H_2 \cdot H_2O$		104
Bisaryl-hydrazon bridge		Aniline		105
Carbamate type of bridge		pH = 4.5		106
Photo-labile (ANP) bridge		λ = 355 nm		107

The high stability of biomolecules under mild basic conditions inspired researchers to focus their efforts on exploring cross-links that are sensitive to alkalines but stable at all other conditions. Hydrolysis of the ester bond is one of the oldest reactions for cleavage that involves formation of a carboxylic acid. Limitations of this reaction are the strongly basic conditions and low yields in case of biomolecules[108,109]. Abdella at al.[101] proposed to use hydroxylamine as nucleophile for the cleavage of ester bonds under alkaline conditions (**Table 1.3. Ester bridge**). According to his approach an esterified covalent bridge can be cleaved under weakly basic conditions, i.e. pH 8.5, for 3–6 hours at 37°C, in the presence of hydroxylamine. The reaction results in the formation of an amide derivative on one fragment and a hydroxyl group on the other fragment.

The formation of an inert amide group after hydroxylamine cleavage inspired new research on cleavable systems. Zarling et al[103]. proposed that the presence of a sulfone group in a cleavable linker can be used for hydrolytic cleavage under basic conditions (**Table 1.3. Sulfone bridges**). In that study, peptide antigens were linked to a surface through a sulfone group, purified and cleaved at pH 11.6 for further analysis. Hydrolysis of the sulfone bridge leads to the formation of a hydroxy- or amino group[110] depending on the specific structure of the linker.

1.10 Cleavable linker based on metal-complexes.

Approaches using coordination chemistry for linker formation are a scarcely investigated area of research. While the covalent organic bridges result in rigid systems with well-defined structures, linker held together by metal-complexes are usually less stable and more dynamic in changing their structures. Differences in ligand stability in the complex can be a critical aspect in chemoselective cleavage of linkers[111,112]. Metal ions that have been used for such complexes include Co(III)[113], Pt(II)[114,115] and Cr(III)[116,117] (**Fig. 1.9**).

Complexes with Pt(II) centers have been the focus of many recent studies dealing with modifications of biomolecules. Complex formation usually involves multidentate ligands with different coordination sites[118]. The structure and stability of bridging metal complexes involving Pt(II) can be directed by a combination of strong-binding (phosphines, amines) and weak-binding sites (O, S, or Se)[119] (**Fig. 1.9 C**).

Civitello and coworkers[116] have recently discussed the possibility to form DNA - peptide conjugates that are connected by a bridge formed by a complex with a Cr(III) metal center. Four coordination sites in the chromium complex are occupied by a peptide tetramer and the remaining two sites by a one nucleobase (or by nucleobase and oxygen atom from phosphodiester bond). In this study a series of peptide tetramers was synthetysed with either serine-glycine or histidine-glycine sequences for the Cr(III)-binding. In the next step authors investigated the possibility to attach chromium complexes to dinucleotides and demonstrated preferential coordination of Cr (III) to guanine through the N7 position as well as the N3 position.

Stayner et al.[120] tested the possibility of using metal-cleavable linker in metal-affinity chromatography. They attached a series of histidine-tyrosine residues to the N-terminus of the I28 immunoglobulin domain of the human cardiac titin. After addition of Ni(II) ions they observed dimerization of peptides due to metal-complex formation.

Gibson et al.[117] proposed that Cr(III) can form complexes with the π-conjugated systems of aromatic rings (**Fig. 1. 9 E**). In their research they attached derivatives of methyl benzoic ester to

a surface through a Cr(III) complex and demonstrated, that this substrate can be efficiently (yield 62%) reduced from ketone derivative to alcohol while preserving the stability of the complex. In the last step quantitative complex cleavage produced 4-(4-methoxypbenyl)-2-butanole.

Arbo and Isied[113] used cobalt complexes with aminobenzoic acid as a linker to polystyrene resin for solid-phase synthesis of enkefalin (**Fig. 1.9 A** and **B**). The protected peptide was cleaved from the surface in 96% yield that makes these complexes competitive with 9-fluorenylmethoxycarbonyl (FMOC) protecting group strategies used in solid-phase peptide synthesis.

2,2'-Bipyridine provides many opportunities as it forms stable complexes with a wide range of metals and is relatively easily functionalized. Killeen at al.[121] demonstrated a series of ruthenium(III) complexes based on this ligand and its capability to cap a membrane pore (**Fig. 1.9 D**).

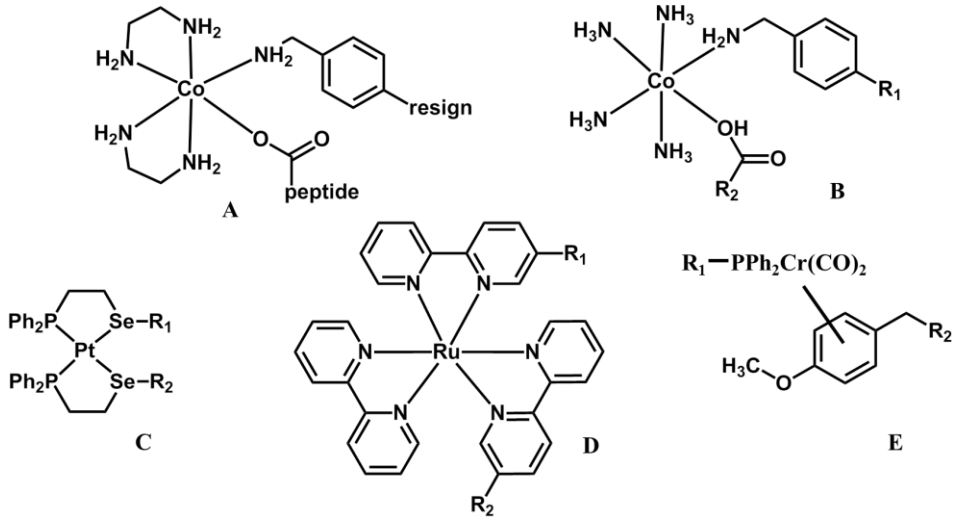


Figure 1.9. Metal-complexes used for the formation of cleavable linkers. For details see text.

1.11 Goals

Non-natural modifications of oligonucleotides have unusual properties and can be used for a range of therapeutic and analytical purposes, including the treatment of diseases, the regulation of gene expression, as well as investigation of DNA and RNA tertiary structures, DNA-peptide interactions, monitoring of RNA concentrations *in vivo* and detection of single nucleotide polymorphism. However, the modification of nucleic acids up to now is either not specific, e.g. complete modification of one kind of nucleotide throughout the whole sequence, or restricted to shorter sequences when using solid-phase synthesis. This prompted us to follow an alternative approach.

The addition of different reactive groups to a recognition sequence allows to recognize a specific region in a native oligonucleotide sequence, to localize the reactive group close in space to the target nucleobase followed by site-specific modification. In the last 20 years this approach was used for the formation of interstrand cross-link products with native DNA strands, e.g. to silence specific gene expression in the cell.

Our objective in the present study is to design a modular system for the site-specific introduction of a functional group into a DNA template using a complementary modified oligonucleotide. The modular system allows to select the reactive group and hence the envisioned modification from all the diversity of chemistry, their independent synthesis, and their assembly to yield the complementary modified oligonucleotide in a simply controllable way.

1.12 References:

(1) Watson, J. D.; Crick, F. H. C. *Nature* **1953**, 737
(2) Mitsui, Y.; Langridge, R.; Shortle, B. E.; Cantor, C. R.; Grant, R. C.; Kodama, M.; Wells, R. D. *Nature* **1970**, *228*, 1166.
(3) Beal, P.; Dervan, P. *Science* **1991**, *251*, 1360.
(4) Phan, A. T.; Leroy, J.-L. *J. Biomol. Struct. Dyn.* **2000**, *17*, 245.
(5) Kim, J.; Doose, S.; Neuweiler, H.; Sauer, M. *Nucleic Acids Res.* **2006**, *34*, 2516.
(6) Cang, X.; Šponer, J.; Cheatham, I. I. I. T. E. *J. Am. Chem. Soc.* **2011**, *133*, 14270.
(7) Pullman, A. *Ann. N.Y. Acad. Sci.* **1969**, *158*, 65.
(8) Giessner-Prettre, C.; Pullman, A. *Theoret. Chim. Acta* **1968**, *9*, 279.
(9) Shabarova, Z. A.; Bogdanov, A. A.; 1 edition (July 13, 1994) ed.; Wiley-VCH: 1994, p 589.
(10) Thuong, N. T.; Hélène, C. *Angew. Chem., Int. Ed.* **1993**, *32*, 666.
(11) Cochran, W. *Acta Crystallogr.* **1951**, *4*, 81.
(12) Kubo, K.; Ide, H.; Wallace, S. S.; Kow, Y. W. *Biochemistry* **1992**, *31*, 3703.
(13) Atamna, H.; Cheung, I.; Ames, B. N. *Proceedings of the National Academy of Sciences* **2000**, *97*, 686.
(14) Lönnberg, H. *Bioconjugate Chem.* **2009**, *20*, 1065.
(15) Guillier, F.; Orain, D.; Bradley, M. *Chem. Rev.* **2000**, *100*, 2091.
(16) Eritja, R. *Int. J. Pept. Res. Ther.* **2007**, *13*, 53.
(17) Khorana, H. G.; Agarwal, K. L.; Büchi, H.; Caruthers, M. H.; Gupta, N. K.; Klbppe, K.; Kumar, A.; Ohtsuka, E.; Bhandary, U. L.; van de Sande, J. H.; Sgaramella, V.; Tebao, T.; Weber, H.; Yamada, T. *J. Mol. Biol.* **1972**, *72*, 209.
(18) Saiki, R.; Gelfand, D.; Stoffel, S.; Scharf, S.; Higuchi, R.; Horn, G.; Mullis, K.; Erlich, H. *Science* **1988**, *239*, 487.
(19) Abarzúa, P.; Marians, K. J. *P. Natl. Acad. Sci. USA* **1984**, *81*, 2030.
(20) Lehman, I. R.; Bessman, M. J.; Simms, E. S.; Kornberg, A. *J. Biol. Chem.* **1958**, *233*, 163.
(21) Yang-Feng, T. L.; Landau, N. R.; Baltimore, D.; Francke, U. *Cytogenet. Genome Res.* **1986**, *43*, 121.
(22) Bartlett, J. S.; Stirling, D. In *PCR Protocols*; Bartlett, J. S., Stirling, D., Eds.; Humana Press: 2003; Vol. 226, p 3.
(23) Goodchild, J. *Bioconjugate Chem.* **1990**, *1*, 165.
(24) Langer, P. R.; Waldrop, A. A.; Ward, D. C. *P. Natl. Acad. Sci. USA* **1981**, *78*, 6633.
(25) Hermanson, G. T.; 2-nd ed.; Elsevier Science: 2008, p 1323.
(26) Molander, J.; Hurskainen, P.; Hovinen, J.; Lahti, M.; Lonnberg, H. *Bioconjugate Chem.* **1993**, *4*, 362.
(27) Chen, H.; Shaw, B. R. *Biochemistry* **1994**, *33*, 4121.
(28) Shapiro, R.; Weisgras, J. M. *Biochem. Biophys. Res. Commun.* **1970**, *40*, 839.
(29) Draper, D. E.; Gold, L. *Biochemistry* **1980**, *19*, 1774.
(30) Okamoto, A.; Tanaka, K.; Saito, I. *J. Am. Chem. Soc.* **2003**, *126*, 416.
(31) Wataya, Y.; Negishi, K.; Hayatsu, H. *Biochemistry* **1973**, *12*, 3992.
(32) Keller, G. H.; Cumming, C. U.; Huang, D.-P.; Manak, M. M.; Ting, R. *Anal. Biochem.* **1988**, *170*, 441.

(33) Rothenberg, J. M.; Wilchek, M. *Nucleic Acids Res.* **1988**, *16*, 7197.
(34) Molenaar, C.; Teuben, J.-M.; Heetebrij, R. J.; Tanke, H. J.; Reedijk, J. *J Biol Inorg Chem* **2000**, *5*, 655.
(35) Babaylova, E.; Graifer, D.; Malygin, A.; Stahl, J.; Shatsky, I.; Karpova, G. *Nucleic Acids Res.* **2009**, *37*, 1141.
(36) Viscidi, R. P. In *Methods Enzymol.*; Meir, W., Edward, A. B., Eds.; Academic Press: 1990; Vol. Volume 184, p 600.
(37) Pluskota-Karwatka, D.; Le Curieux, F.; Munter, T.; Sjöholm, R.; Kronberg, L. *Chem. Res. Toxicol.* **2002**, *15*, 110.
(38) Kellner, S.; Seidu-Larry, S.; Burhenne, J.; Motorin, Y.; Helm, M. *Nucleic Acids Res.* **2011**, *39*, 7348.
(39) Sancar, A.; Lindsey-Boltz, L. A.; Ünsal-Kaçmaz, K.; Linn, S. *Annu. Rev. Biochem* **2004**, *73*, 39.
(40) Noll, D. M.; Mason, T. M.; Miller, P. S. *Chem. Rev.* **2005**, *106*, 277.
(41) Luce, R. A.; Hopkins, P. B. In *Methods Enzymol.*; Jonathan B. Chaires, M. J. W., Ed.; Academic Press: 2001; Vol. Volume 340, p 396.
(42) Efimov, V. A.; Fedyunin, S. V.; Chakhmakhcheva, O. G. *Russ. J. Bioorg. Chem.* **2010**, *36*, 49.
(43) Burchenal, J. H.; Burchenal, J. R.; Johnston, S. F. *Cancer* **1951**, *4*, 353.
(44) Kasparkova, J.; Vojtiskova, M.; Natile, G.; Brabec, V. *Chem–Eur. J.* **2008**, *14*, 1330.
(45) Reishus, J. W.; Martin, D. S. *J. Am. Chem. Soc.* **1961**, *83*, 2457.
(46) Aprile, F.; Martin, D. S. *Inorg. Chem.* **1962**, *1*, 551.
(47) Yang, D.; van Boom, S. S. G. E.; Reedijk, J.; van Boom, J. H.; Wang, A. H. J. *Biochemistry* **1995**, *34*, 12912.
(48) Lemaire, M. A.; Schwartz, A.; Rahmouni, A. R.; Leng, M. *Proceedings of the National Academy of Sciences* **1991**, *88*, 1982.
(49) Cox, J. W.; Berners-Price, S. J.; Davies, M. S.; Qu, Y.; Farrell, N. *J. Am. Chem. Soc.* **2001**, *123*, 1316.
(50) Gruff, E. S.; Orgel, L. E. *Nucleic Acids Res.* **1991**, *19*, 6849.
(51) Walker, W. L.; Kopka, M. L.; Filipowsky, M. E.; Dickerson, R. E.; Goodsell, D. S. *Biopolymers* **1995**, *35*, 543.
(52) Hertzberg, R. P.; Hecht, S. M.; Reynolds, V. L.; Molineux, I. J.; Hurley, L. H. *Biochemistry* **1986**, *25*, 1249.
(53) Hurley, L. H.; Reck, T.; Thurston, D. E.; Langley, D. R.; Holden, K. G.; Hertzberg, R. P.; Hoover, J. R. E.; Gallagher, G.; Faucette, L. F. *Chem. Res. Toxicol.* **1988**, *1*, 258.
(54) Kohn, K. W.; Glaubiger, D.; Spears, C. L. *Biochimica et Biophysica Acta (BBA) - Nucleic Acids and Protein Synthesis* **1974**, *361*, 288.
(55) Rink, S. M.; Lipman, R.; Alley, S. C.; Hopkins, P. B.; Tomasz, M. *Chem. Res. Toxicol.* **1996**, *9*, 382.
(56) Norman, D.; Live, D.; Sastry, M.; Lipman, R.; Hingerty, B. E.; Tomasz, M.; Broyde, S.; Patel, D. J. *Biochemistry* **1990**, *29*, 2861.
(57) Bose, D. S.; Thompson, A. S.; Ching, J.; Hartley, J. A.; Berardini, M. D.; Jenkins, T. C.; Neidle, S.; Hurley, L. H.; Thurston, D. E. *J. Am. Chem. Soc.* **1992**, *114*, 4939.
(58) Mountzouris, J. A.; Wang, J.-J.; Thurston, D.; Hurley, L. H. *J. Med. Chem.* **1994**, *37*, 3132.
(59) Jackson, C.; Hartley, J. A.; Jenkins, T. C.; Godfrey, R.; Saunders, R.; Thurston, D. E. *Biochem. Pharmacol.* **1991**, *42*, 2091.
(60) Jackson, C.; Crabb, T. A.; Gibson, M.; Godfrey, R.; Saunders, R.; Thurston, D. E. *J. Pharm. Sci.* **1991**, *80*, 245.
(61) Cimino, G. D.; Gamper, H. B.; Isaacs, S. T.; Hearst, J. E. *Annu. Rev. Biochem* **1985**, *54*, 1151.
(62) Straub, K.; Kanne, D.; Hearst, J. E.; Rapoport, H. *J. Am. Chem. Soc.* **1981**, *103*, 2347.
(63) Turner, P. R.; Denny, W. A. *Mutat. Res-Fund. Mol. M.* **1996**, *355*, 141.
(64) Choudhury, J.; Rao, L.; Bierbach, U. *J. Biol. Inorg. Chem.* **2011**, *16*, 373.
(65) Paul, M.; Murray, V. *J. Biol. Inorg. Chem.* **2011**, *16*, 735.
(66) Basu, A. K.; Essigmann, J. M. *Chem. Res. Toxicol.* **1988**, *1*, 1.
(67) Tsuji, A.; Sato, Y.; Hirano, M.; Suga, T.; Koshimoto, H.; Taguchi, T.; Ohsuka, S. *Biophys. J.* **2001**, *81*, 501.
(68) Gates, K. S.; Nooner, T.; Dutta, S. *Chem. Res. Toxicol.* **2004**, *17*, 839.
(69) Ziemba, A. J.; Reed, M. W.; Raney, K. D.; Byrd, A. B.; Ebbinghaus, S. W. *Biochemistry* **2003**, *42*, 5013.
(70) Moser, H.; Dervan, P. *Science* **1987**, *238*, 645.
(71) Taylor, M. J.; Dervan, P. B. *Bioconjugate Chem.* **1997**, *8*, 354.
(72) Grant, K. B.; Dervan, P. B. *Biochemistry* **1996**, *35*, 12313.
(73) Kuznetsova, A. A.; Solov'eva, L. I.; Fedorova, O. S. *Russ. J. Bioorg. Chem.* **2008**, *34*, 613.
(74) Hall, J.; Hüsken, D.; Häner, R. *Nucleic Acids Res.* **1996**, *24*, 3522.
(75) Magda, D.; Miller, R. A.; Sessler, J. L.; Iverson, B. L. *J. Am. Chem. Soc.* **1994**, *116*, 7439.
(76) Hovinen, J.; Guzaev, A.; Azhayeva, E.; Azhayev, A.; Lonnberg, H. *J. Org. Chem.* **1995**, *60*, 2205.
(77) Bashkin, J. K.; Frolova, E. I.; Sampath, U. *J. Am. Chem. Soc.* **1994**, *116*, 5981.
(78) Joyner, J. C.; Cowan, J. A. *J. Am. Chem. Soc.* **2011**, *133*, 9912.
(79) Hentschel, S.; Alzeer, J.; Angelov, T.; Schärer, O. D.; Luedtke, N. W. *Angew. Chem. Int. Ed.* **2012**, *51*, 3466.
(80) Stevens, K.; Claeys, D. D.; Catak, S.; Figaroli, S.; Hocek, M.; Tromp, J. M.; Schürch, S.; Van Speybroeck, V.; Madder, A. *Chem–Eur. J.* **2011**, *17*, 6940.
(81) Stevens, K.; Madder, A. *Nucleic Acids Res.* **2009**, *37*, 1555.
(82) Ren, S.; Cai, L.; M. Segal, B. *J. Chem. Soc., Dalton Trans.* **1999**, 1413.
(83) Sharma, S. K.; McLaughlin, L. W. *J. Am. Chem. Soc.* **2002**, *124*, 9658.
(84) Vasquez, K. M.; Wensel, T. G.; Hogan, M. E.; Wilson, J. H. *Biochemistry* **1996**, *35*, 10712.

(85) Jahn, K.; Olsen, E. M.; Nielsen, M. M.; Tørring, T.; Mohammad, R.; Andersen, E. S.; Gothelf, K. V.; Kjems, J. *Bioconjugate Chem.* **2010**, *22*, 95.
(86) Agapkina, Y. Y.; Agapkin, D. V.; Zagorodnikov, A. V.; Alekseev, Y. I.; Korshunova, G. A.; Gottikh, M. B. *Russ. J. Bioorg. Chem.* **2002**, *28*, 293.
(87) Bielski, R.; Witczak, Z. *Chem. Rev.* **2012**.
(88) Gartner, C. A.; Elias, J. E.; Bakalarski, C. E.; Gygi, S. P. *J. Proteome Res.* **2007**, *6*, 1482.
(89) Burns, J. A.; Butler, J. C.; Moran, J.; Whitesides, G. M. *J. Org. Chem.* **1991**, *56*, 2648.
(90) Partis, M. D.; Griffiths, D. G.; Roberts, G. C.; Beechey, R. B. *J. Protein Chem.* **1983**, *2*, 263.
(91) Abe, H.; Wakabayashi, R.; Yonemura, H.; Yamada, S.; Goto, M.; Kamiya, N. *Bioconjugate Chem.* **2013**, *24*, 242.
(92 Liu, F.; Soh Yan Ni, A.; Lim, Y.; Mohanram, H.; Bhattacharjya, S.; Xing, B. *Bioconjugate Chem.* **2012**, *23*, 1639.
(93) Bishop, J. E.; Squier, T. C.; Bigelow, D. J.; Inesi, G. *Biochemistry* **1988**, *27*, 5233.
(94) Højfeldt, J. W.; Gothelf, K. V. *J. Org. Chem.* **2006**, *71*, 9556.
(95) Hoppmann, C.; Schmieder, P.; Heinrich, N.; Beyermann, M. *ChemBioChem* **2011**, *12*, 2555.
(96) Gartner, Z. J.; Kanan, M. W.; Liu, D. R. *J. Am. Chem. Soc.* **2002**, *124*, 10304.
(97) Angelov, T.; Guainazzi, A.; Schärer, O. D. *Org. Lett.* **2009**, *11*, 661.
(98) Zatsepin, T. S.; Stetsenko, D. A.; Gait, M. J.; Oretskaya, T. S. *Bioconjugate Chem.* **2005**, *16*, 471.
(99) Hileman, R. E.; Parkhurst, K. M.; Gupta, N. K.; Parkhurst, L. J. *Bioconjugate Chem.* **1994**, *5*, 436.
(100) Verhelst, S. H. L.; Fonović, M.; Bogyo, M. *Angew. Chem. Int. Ed.* **2007**, *46*, 1284.
(101) Abdella, P. M.; Smith, P. K.; Royer, G. P. *Biochem. Biophys. Res. Commun.* **1979**, *87*, 734.
(102) Antsypovich, S. I.; von Kiedrowski, G. *Nucleos. Nucleot. Nucl.* **2005**, *24*, 211.
(103) Zarling, D. A.; Watson, A.; Bach, F. H. *J. Immunol.* **1980**, *124*, 913.
(104) Geurink, P. P.; Florea, B. I.; Li, N.; Witte, M. D.; Verasdonck, J.; Kuo, C.-L.; van der Marel, G. A.; Overkleeft, H. S. *Angew. Chem. Int. Ed.* **2010**, *49*, 6802.
(105) Dirksen, A.; Yegneswaran, S.; Dawson, P. E. *Angew. Chem. Int. Ed.* **2010**, *49*, 2023.
(106) Ball, H. L.; Mascagni, P. *J. Pept. Sci.* **1997**, *3*, 252.
(107) Brown, B.; Wagner, D.; Geysen, H. M. *Mol Divers* **1995**, *1*, 4.
(108) Kurth, M. J.; Ahlberg Randall, L. A.; Takenouchi, K. *J. Org. Chem.* **1996**, *61*, 8755.
(109) Evans, D. A.; Black, W. C. *J. Am. Chem. Soc.* **1993**, *115*, 4497.
(110) Gartner, Z. J.; Tse, B. N.; Grubina, R.; Doyon, J. B.; Snyder, T. M.; Liu, D. R. *Science* **2004**, *305*, 1601.
(111) Yoon, H. J.; Mirkin, C. A. *J. Am. Chem. Soc.* **2008**, *130*, 11590.
(112) Yoon, H. J.; Heo, J.; Mirkin, C. A. *J. Am. Chem. Soc.* **2007**, *129*, 14182.
(113) Arbo, B. E.; Isied, S. S. *Int. J. Pept. Protein Res.* **1993**, *42*, 138.
(114) Spokoyny, A. M.; Rosen, M. S.; Ulmann, P. A.; Stern, C.; Mirkin, C. A. *Inorg. Chem.* **2010**, *49*, 1577.
(115) Ulmann, P. A.; Brown, A. M.; Ovchinnikov, M. V.; Mirkin, C. A.; DiPasquale, A. G.; Rheingold, A. L. *Chem–Eur. J.* **2007**, *13*, 4529.
(116) Civitello, E. R.; Leniek, R. G.; Hossler, K. A.; Haebe, K.; Stearns, D. M. *Bioconjugate Chem.* **2001**, *12*, 459.
(117) Gibson, S. E.; Hales, N. J.; Peplow, M. A. *Tetrahedron Lett.* **1999**, *40*, 1417.
(118) Kašpárková, J.; Nováková, O.; Vrána, O.; Farrell, N.; Brabec, V. *Biochemistry* **1999**, *38*, 10997.
(119) Rosen, M. S.; Spokoyny, A. M.; Machan, C. W.; Stern, C.; Sarjeant, A.; Mirkin, C. A. *Inorg. Chem.* **2010**, *50*, 1411.
(120) Stayner, R. S.; Min, D.-J.; Kiser, P. F.; Stewart, R. J. *Bioconjugate Chem.* **2005**, *16*, 1617.
(121) Scott Killeen, J.; Browne, W. R.; Skupin, M.; Fuhrhop, J.-H.; Vos, J. G. *New J. Chem.* **2003**, *27*, 1078.

CHAPTER 2

General principle of modular strategy

2.1 Introduction

Oligonucleotides and oligonucleotide analogues conjugated with functional groups have unusual properties which make them useful for biomedical and therapeutic applications. These oligonucleotides can be used as a probe for polymerase chain reaction (PCR), for the immobilization of molecules on surfaces, for the detection of genetic mutations, as probes for *in vivo* imaging etc[1]. The functional groups can be included at the end of the oligonucleotide molecule as well as in the middle of the sequence. Numerous protocols provide easy and efficient oligonucleotide labeling at the end of the sequence through the phosphate group[2] or by breaking the sugar ring[3]. However, site-specific labeling of long native oligonucleotides with preserved tertiary structure is still a challenging task for chemists and biologists. In our research we propose a modular strategy for the simple and fast introduction of functional groups at any position within a native oligonucleotide irrespective of its length.

2.2 Construction of the modular system.

Our goal was to design a system that enables the chemoselective modification of specific positions in an oligonucleotide in presence of a wide variety of biomolecules. The main focus was to design a system consisting of building blocks that can be prepared in relatively few synthetic steps from low-cost commercially available materials to make our method convenient and an alternative to the widely used solid-phase prepared immobilization probes. We implemented DNA-template chemistry to establish a simple, modular, and multiplexed system for site-specific labeling of unmodified oligonucleotides with small molecules. Such modified nucleic acids[4] (e.g. fluorophore labeled oligonucleotides) can be used e.g. to visualize the distance dependency during FRET. The general structure of the modular system presented here consists of three parts (**Fig. 2.1**): (i) a complementary oligonucleotide sequence, denoted as Short Recognition Sequence (SRS) which is required to sense the specific position in the target DNA template by Watson-Crick base pair formation, (ii) a reactive group to generate the desired modification, and (iii) a cleavable linker connecting the recognition sequence with the reactive group. After cleavage of the initially formed covalent interstrand cross-link between the recognition and the target sequence the desired functional group will be generated. This group

can be used for labeling with e.g. a fluorophore. Simplicity of the design of the reactive group and connecting linker is an important advantage of this method.

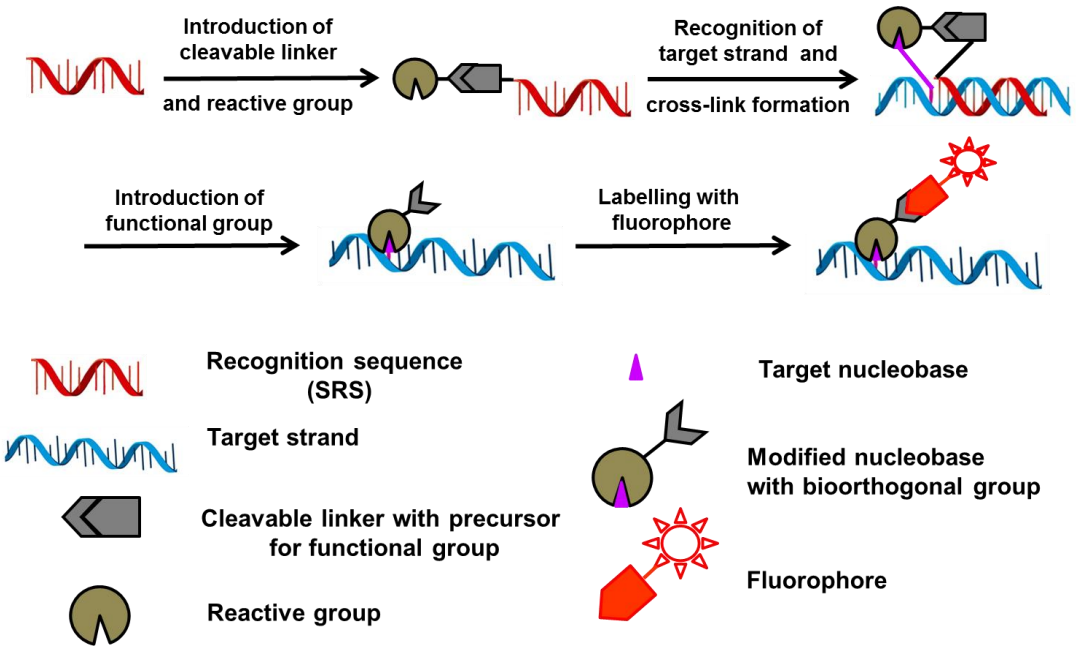


Figure 2.1. Modular strategy for site-specific modification and labeling of native oligonucleotides.

2.3 Short recognition sequence (SRS).

The main key of our research is the site-specific modification of native oligonucleotides and nucleic acids. Over the years, different strategies were proposed for site-specific reaction with target oligonucleotides. These strategies differ in the sort of molecule used for binding to the oligonucleotide: duplex formation with complementary oligonucleotide primer[5,6], triplex forming systems[7-9], site-specific intercalating reagents[10,11], zinc-finger nucleases[12-15] and transcription-activator like (TAL) effectors[16-18]. The approach to use complementary oligonucleotide strands is very convenient due to its simplicity, low price of initial chemicals and easy availability. The efficiency of binding between target and recognition strands is determined by their affinity and selectivity[19]. The length of the recognition sequence is an important parameter that determines these two parameters. Recognition and formation of a helical structure with sequences of the human genome requires at least 15-nucleotide-long donor-target pairs[20]. At the same time, depending on the GC content, pairs of such length can be too stable to be sensitive to a single base mispair (affinity problem). In case of energy factors, a single mismatch decreases the stabilization free energy by 3-5 kcal/mol[21,22]. For long complementary sequences of e.g. 50 nucleotides in length, a single mismatch only cause minor changes in duplex stability and an insignificant decrease of the melting temperature of the duplex. Accordingly, site-

specificity is reduced and the potential localization of the reactive group close to a wrong target nucleotide can significantly decrease the yield of target nucleotide modification. Based on melting point studies[23] the optimal length for the recognition sequence should be around 15-30 DNA bases[24] that corresponds to melting temperatures between 30°C ─ 50°C. In this way, the duplex stability still allows efficient localization of the reactive group close to the desired region in the target DNA, while the stability is still low enough to allow a significantly decrease lifetime and stability of the DNA duplex in case of a single basepair mismatch in the sequence. Highly interesting are also reactive sequences that contain the reactive group in the middle of the recognition sequence as the flanking complementary nucleotides ensure a rather rigid positioning of the reactive group opposite the target base. The adaptation of our modular system to the different modification tasks is straight forward due to our modular approach, in which the linker as well as the reactive group are synthesized separately and only then assembled to give the reactive sequence.

The driving force of DNA-templated chemistry is the recognition of a specific region in a template oligonucleotide by a recognition sequence. Liu and coworkers[25] demonstrated the importance of annealing by combinatorial synthesis of a series of organic molecules. They divided synthesis of cyclic molecules into six steps and attached each precursor to the specific guide sequences. The stepwise synthesis of an organic molecule was performed directly on the DNA-template according to complementary regions. Further evidence for the importance of duplex formation was demonstrated by Seitz at al[26,27]. Similar to previous result, they attached amino-acid derivatives to recognition sequences. After annealing of the template DNA, they initiated peptide synthesis by native chemical ligation. By changing the order of complementary regions on the DNA template they obtained peptides with different sequences.

2.4 Chemoselective cleavable part.

In our modular system the linker connects the reactive group with the recognition sequence. On the one hand, the linker should spatially separate the reactive group from the recognition sequence to prevent self-modification as far as possible. On the other hand the linker should be short enough to enable site-specific reaction with the target in the template strand. Accordingly, linker size and flexibility have to be designed and optimized for each type of target. Previous methods for attaching linker to oligonucleotides rely on non-selective post-synthetic modifications, e.g. biotinylation of single stranded DNA *via* transamination of cytosine[28,29] and NBS-catalyzed C8-attachment to guanine[30], or on site-specific solid-phase synthesis[31-33]. Our interest was focused on finding low-cost and easy methods of incorporation, to allow fast and selective conjugation of the linker. Hence we proposed three major requirements for the linker:

1) Short linker size and simple structure to minimize interference with the DNA chain. However, too small linkers are probably not long enough to reach the target nucleotide.

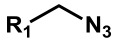
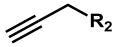
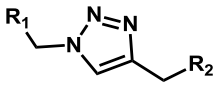
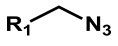
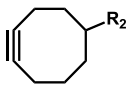
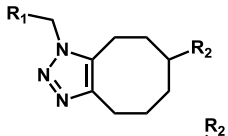
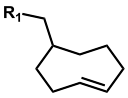
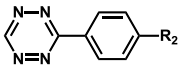
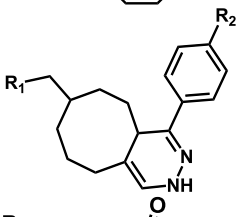
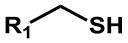
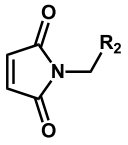
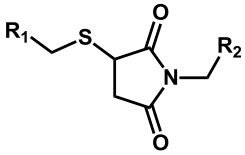
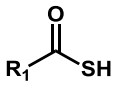
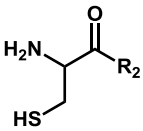
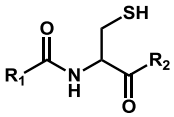
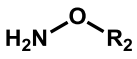
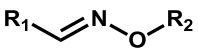
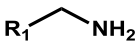
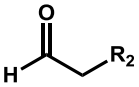
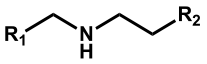
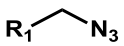
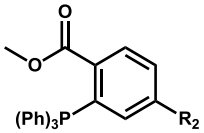
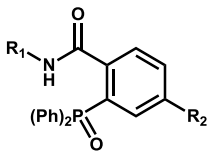
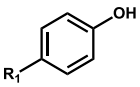
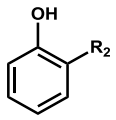
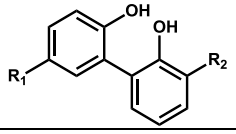
2) Linker should contain at least two orthogonal groups for attaching to DNA recognition sequence in one side as well to the reactive group in the other end.

3) Linker should contain chemoselective cleavable moiety.

For the second requirement there are two general designs of reactive sequence; one with the reactive group added to the end of the oligonucleotide primer and one where the reactive group is incorporated in the middle of the sequence. In the region, where DNA helix structure ends and double strand structure transforms into single strand a center of destabilization is created. This destabilization is induced because the lack of missing neighboring base-pairs at the helix end leads to a reduction of energy from π-π base stacking and hydrogen bond formation. Some groups[34] investigated DNA helix unzipping mechanisms and proposed that some of the last nucleobases in duplex regions are not fixed, but rather are in equilibrium between two single strands and a rigid helix structure which is called "DNA end-breathing"[35]. To avoid this effect we proposed the construct mentioned above, where cleavable linker and reactive group are incorporated in the middle of the SRS. At both ends the construct is locked by 10 nucleotide long complementary sequences. Incorporation of the linker in the middle of the recognition sequence leads to a strong localization of the reactive group in space, a stabilization of the duplex and increasing yields. The drawback is that such incorporation can be a challenging task. In contrast, while modification of the recognition sequence at the end is less selective and leads to lower yields due to DNA breathing the chemical synthesis is more straightforward. To provide chemists with different methodologies for the incorporation of reactive groups, we focused on transamination of cytosine[36,37] as a method for incorporation of the linker in the middle of the reactive sequence. As an alternative approach, incorporation of the linker to the 5'-end through a phosphoramine bond was done according to Boutorine's protocol[2]. In both methodologies low-cost non-modified DNA recognition sequences were used as starting compounds.

Importantly, coupling of the linker to the recognition sequence should proceed under aqueous conditions and should not interfere with other biomolecules. Furthermore, the different reactions for the coupling of linker to recognition sequence and reactive group is an important aspect in our research. We assumed that bioorthogonal reactions[38,39], in particular the oxime ligation, Staudinger reduction, Diels-Alder reaction, chemical native ligation, amide coupling, reductive amination between amines and aldehyde group, as well as the Cu(I)-catalyzed azide – alkyne cycloaddition can serve as powerful tools for chemoselective couplings which fulfill all requirements for our purpose.

Table 2.1. Classic bioconjugation reactions for assembling reactive, cleavable and appropriate recognition units in one probe.

Entry	Name	Reagent 1	Reagent 2	Product	Ref.
1	[3+2] dipolar cycloaddition catalysed by Cu (I)				38
2	[3+2] dipolar cycloaddition copper free				40
3	Diels-Alder cycloaddition				41
4	Michael addition				42
5	Native chemical ligation				43
6	Oxime condensation				44
7	Reductive amination of aldehyde				45
8	Staudinger ligation				46
9	Metal-mediated oxidation				47

The third requirement for the linker in our modular system is the presence of a cleavage site to generate a useful functional group at the target nucleobase after cross-link formation between reactive and target sequence. Traditional linkers, e.g. levulinoyl ester or azobenzene derivatives[48], provide products with some side chains after cleavage causing undesirable biological or chemical activity of the target molecule[49]. For these reasons, there has been significant interest in developing so-called "traceless linkers"[50]. In this approach, the group generated after cleavage is immediately further modified to a group compatible with nucleic acid chemistry. To avoid undesired side reactions we focused on cleavage sites that result in functional groups that do not further react with nucleic acids but allow further modification with specific reagents. We created a library of linkers with different sizes (**Chapter 3, Tables 3.2** and

3.3 and **chapter 5.6**), a number of diverse cleavable parts such as diol- or disulfide-groups (**Fig. 2.2**), and with varying flexibility. After formation of the cross-linked duplex we demonstrate cleavage of linker and *in situ* capturing of the newly formed reactive group.

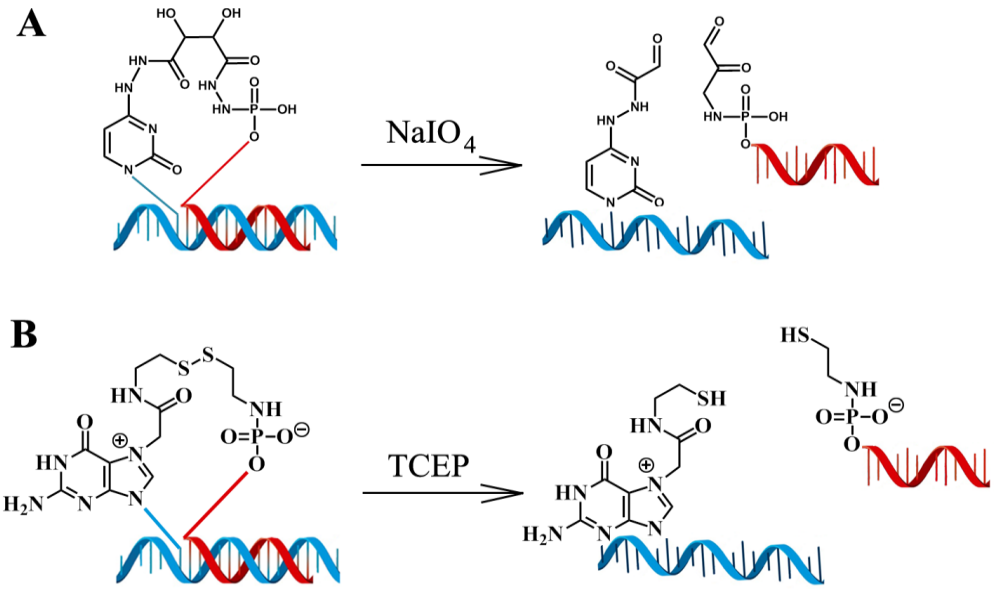


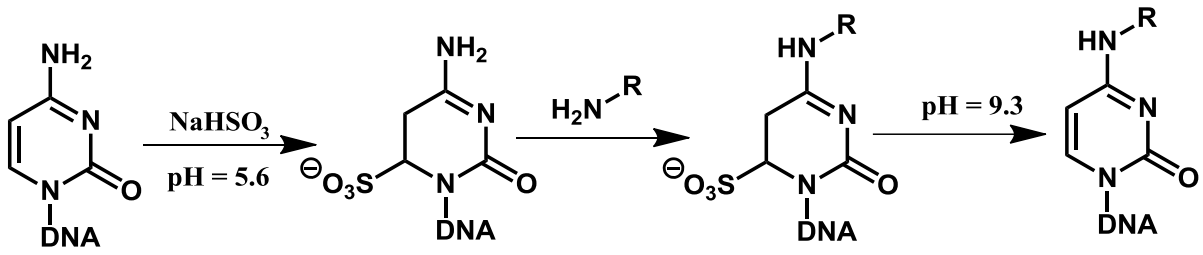
Figure 2.2. Diol- (**A**) and disulfide- (**B**) cleavable linkers.

2.5 Reactive group

In our research we focuse on the modification of single stranded oligonucleotide regions. Accordingly, the reactive group was designed in such a way that it can selectively modify a specific nucleotide in the presence of the three other classes of nucleotides and different positions for nucleotide modifications are possible. As an additional requirement, the reactive group needs a linkage site from where it can be coupled to the short recognition sequence. In general, labeled oligonucleotide strands will interact with other molecules as well as reorganize to form secondary and tertiary structures and different sites of modification can start to influence on stability of hydrogen bonds. Depending on the site, the modification can preserve, with or without distortion, or hinder Watson-Crick base pairing[51,52]. In our research we demonstrated modifications that preserve, distort or hinder Watson-Crick base pairing (**Table 2.2**).

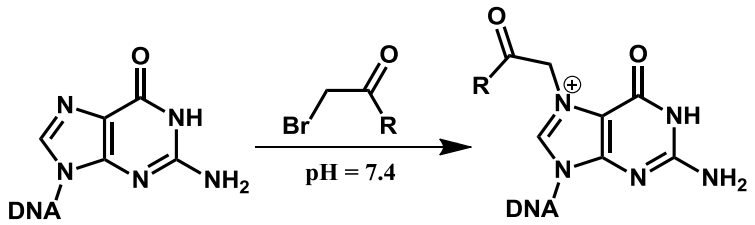
1) Distortion of nucleobase interaction. Modification of cytosine is performed at the exocyclic amino group, which is involved in Watson-Crick base pairing. For this we use an amino group-containing molecule as a reactive group (**Scheme 2.1**) and activate cytosine by hydrogensulfite addition to the 5,6-double bond. In the next step, the nucleophilic reactive group replaces the N4-amino group in 6-sulfo-cytidine. Depending on the orientation of this new

secondary amino group, the amino proton can still participate in Watson-Crick hydrogen bonding, although probably causing a slight distortion due to steric effect.



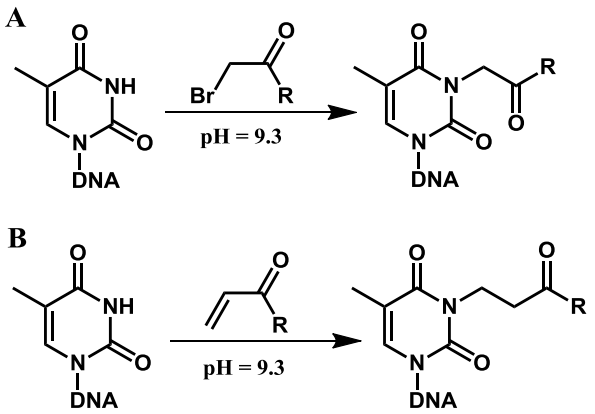
Scheme 2.1. Chemical transformation of cytosine is catalysed by hydrogensulfite[51].

2) Preservation of nucleobase interaction. In contrast, hydrogen bonds between target and complementary nucleotides can be preserved when using the N7-position of guanosine for modification (**Scheme 2.2**). The N7 position is located in the major groove that makes it accessible not only in single stranded DNA, but also in double strand regions. When a suitable electrophilic reactive group, e.g. halogen-acetates or nitrogen mustard[58] is localized close to a target guanine by intercalation[59] or *via* triple helix formation[60] fast alkylation is achieved with preservation of Watson-Crick hydrogen bonds. Unfortunately, N^7-alkylated guanines can be easily depurinated under physiological conditions[61] and the rate of spontaneous loss of the purine base cannot be decreased by changes of the buffer[62,63]. Alternatively, also Pt(II) complexes can be coordinated to the N7 position of guanine resulting in a Pt-G adduct with high stability under different conditions[64].



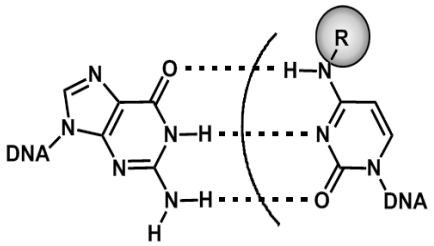
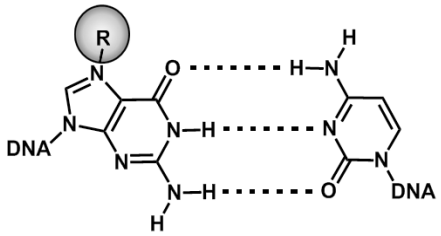
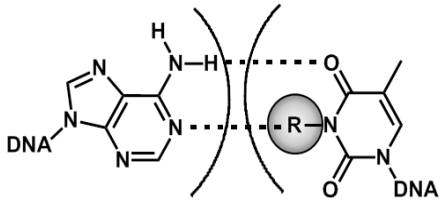
Scheme 2.2. Alkylation of N7 nitrogen atom in guanosine by halogen-acetate derivatives.

3) Hindrance of nucleobase interaction. Modification of thymine N3 prevents formation of Watson-Crick hydrogen bonds (**Scheme 2.3**). The N3-H proton has a pK_a value of 9[65] and hence the N3 position can be used as a target for alkylation of thymidine under basic conditions. Alkylation can be achieved by 2-halogenacetates as reactive group. However, to avoid alkylation of guanine as a side reaction we also evaluate the use of acrylic ester as reactive group (**Table 2.2, entry 3**), as the terminal carbon atom carries a partial positive charge[66] similar to 2-halogenacetates.



Scheme 2.3. Alkylation of thymidine with 2-bromoacetic acid (**A**) or Michael addition (**B**) under alkaline conditions.

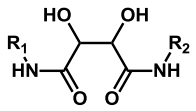
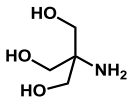
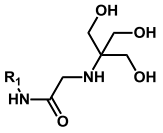
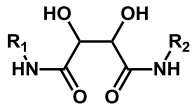
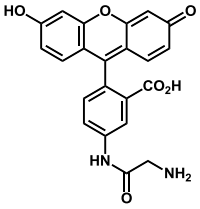
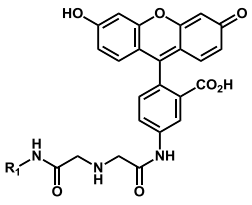
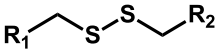
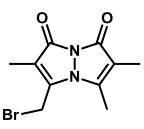
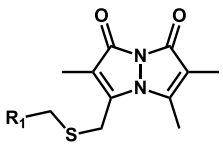
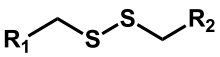
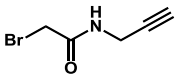
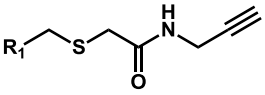
Table 2.2. Modification of nucleobases that preserve, distort or hinder Watson-Crick base pairing.

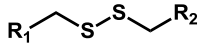
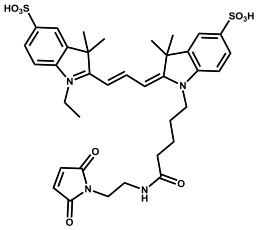
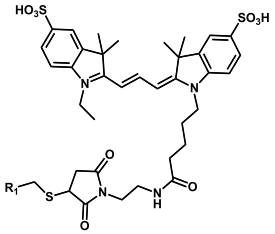
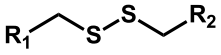
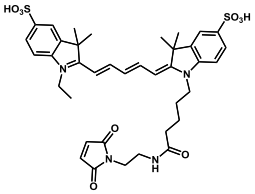
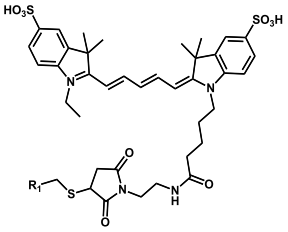
Entry	Target	Reactive group	Conditions	Watson-Crick distortion
1	N^4-position in cytidine	NH_2─R N_2H_2─R	$NaHSO_3$ pH = 5.6	
2	N^7-position in guanine	Br─CH_2COR Cl─$(CH_2)_2$NR cistamine	pH = 7.4	
3	N^3-position in thymidine	Br─CH_2COR CH_2=$CHCH_2$COR	pH = 8.6	

2.5 Modular system: Reaction pathways

In contrast to previous reports describing the labeling of oligonucleotides[67-69] we used native unmodified oligonucleotides as a target. Formation of a duplex structure between reactive and target sequences reduces the mobility of the reactive group and is of central importance for the site-specific interstrand cross-link formation. The reaction can be induced by external triggers such as an additional reagent or irradiation with light or can be self-induced by the proximity of the reactive group to the target nucleobase. After reaction, the covalent cross-link can be cleaved due to the presence of a cleavage site. In the last step we demonstrated conjugation of different molecules to the newly formed functional groups (**Table 2.3**). The modular approach was used for the modification of long target strands up to 50 nucleotides and different functional groups were efficiently introduced, such as aldehyde, carboxyl and thiol groups. The newly introduced group was efficiently used in a further step for labeling with a fluorophore, i.e. Michael addition of a maleimide-Cy3 dye to a thiol, etc (**Table 2.3**). The site-specific modifications in the target strands were detected by PAGE, MALDI-TOF, ESI-MS, and LC-MS analyses of enzymatic digestions and used in FRET applications.

Table 2.3. Examples of molecules that can be conjugated to the newly formed functional groups after linker cleavage.

Entry	Linker	Conjugated compound	Conditions for conjugation	Product	Chapter
1			i) $NaIO_4$ ii) $NaHSO_3$ iii) $NaCNBH_3$		Chapter 3
2			i) $NaIO_4$ ii) $NaHSO_3$ iii) $NaCNBH_3$		Chapter 3
3			TCEP		Chapter 4
4			TCEP		Chapter 5

5			TCEP		Chapter 6
6			TCEP		Chapter 6

2.6 Conclusion

We designed a new general approach for the site-specific modification of native oligonucleotides. Reactive sequences were assembled from three independent parts. Simple conjugation protocols provide the opportunity for designing the most suitable reactive sequence for the respective experiment. The reagents needed for the syntheses of all systems are commercially available at rather low-cost and the synthetic procedures are straightforward. In general, the modular system allows the straightforward exchange of the recognition sequence to enable recognition of specific positions in both DNA and RNA oligonucleotides. In principal, also other recognition units are feasible so that the modular system could be further extended to the modification of other biomolecules e.g. enzymes, lipids, etc. For the future purpose, modular strategies provide tools for site-specific labeling of native nucleic acids in two, three or multiple positions in one step by incubation of the target strand with a mixture of different reactive sequences. To provide effective modifications, reactive sequences should differ from each other by combination recognition sequence and cleavable linker that provides respective bioorthogonal groups after cleavage in appropriate positions.

2.7 References:

(1) Liu, J.; Cao, Z.; Lu, Y. *Chem. Rev.* **2009**, *109*, 1948.
(2) Grimm, G. N.; Boutorine, A. S.; Helene, C. *Nucleos. Nucleot. Nucl.* **2000**, *19*, 1943.
(3) Paredes, E.; Evans, M.; Das, S. R. *Methods* **2011**, *54*, 251.
(4) Coleman, R. S.; Kesicki, E. A. *J. Org. Chem.* **1995**, *60*, 6252.
(5) Stevens, K.; Claeys, D. D.; Catak, S.; Figaroli, S.; Hocek, M.; Tromp, J. M.; Schürch, S.; Van Speybroeck, V.; Madder, A. *Chem–Eur. J.* **2011**, *17*, 6940.
(6) Tsuji, A.; Koshimoto, H.; Sato, Y.; Hirano, M.; Sei-Iida, Y.; Kondo, S.; Ishibashi, K. *Biophys. J.* **2000**, *78*, 3260.
(7) Ziemba, A. J.; Reed, M. W.; Raney, K. D.; Byrd, A. B.; Ebbinghaus, S. W. *Biochemistry* **2003**, *42*, 5013.

(8) Chan, P. P.; Glazer, P. M. *J. Mol. Med.* **1997**, *75*, 267.
(9) Shahid, K. A.; Majumdar, A.; Alam, R.; Liu, S.-T.; Kuan, J. Y.; Sui, X.; Cuenoud, B.; Glazer, P. M.; Miller, P. S.; Seidman, M. M. *Biochemistry* **2006**, *45*, 1970.
(10) Ghosh, S.; Defrancq, E.; Lhomme, J. H.; Dumy, P.; Bhattacharya, S. *Bioconjugate Chem.* **2004**, *15*, 520.
(11) Gianolio, D. A.; McLaughlin, L. W. *J. Am. Chem. Soc.* **1999**, *121*, 6334.
(12) Osakabe, K.; Osakabe, Y.; Toki, S. *P. Natl. Acad. Sci. USA* **2010**, *107*, 12034.
(13) Townsend, J. A.; Wright, D. A.; Voytas, D. F. *Nature* **2009**, *459*, 442
(14) Cai, C.; Doyon, Y.; Ainley, W. M.; Miller, J.; DeKelver, R.; Moehle, E.; Rock, J.; Lee, Y.-L.; Garrison, R.; Schulenberg, L.; Blue, R.; Worden, A.; Baker, L.; Faraji, F.; Zhang, L.; Holmes, M.; Rebar, E.; Collingwood, T.; Rubin-Wilson, B.; Gregory, P.; Urnov, F.; Petolino, J. *Plant Mol. Biol.* **2009**, *69*, 699.
(15) Cathomen, T.; Keith Joung, J. *Mol. Ther.* **2008**, *16*, 1200
(16) Sung, Y. H.; Lee, H.-W. *Nat. Biotechnol.* **2013**, *31*, 23
(17) Wefers, B.; Meyer, M.; Ortiz, O.; Hrabé de Angelis, M.; Hansen, J.; Wurst, W.; Kühn, R. *P. Natl. Acad. Sci. USA* **2013**.
(18) Bogdanove, A. J.; Voytas, D. F. *Science* **2011**, *333*, 1843.
(19) Demidov, V. V.; Frank-Kamenetskii, M. D. *Trends Biochem. Sci* **2004**, *29*, 62.
(20) Cisse, I. I.; Kim, H.; Ha, T. *Nat. Struct. Mol. Biol.* **2012**, *19*, 623.
(21) Tibanyenda, N.; De Bruin, S. H.; Haasnoot, C. A. G.; Van Der Marel, G. A.; Van Boom, J. H.; Hilbers, C. W. *Eur. J. Biochem.* **1984**, *139*, 19.
(22) Aboul-ela, F.; Koh, D.; Tinoco, I.; Martin, F. H. *Nucleic Acids Res.* **1985**, *13*, 4811.
(23) Mergny, J.-L.; Lacroix, L. *Oligonucleotides* **2003**, *13*, 515.
(24) Igloi, G. L. *P. Natl. Acad. Sci. USA* **1998**, *95*, 8562.
(25) He, Y.; Liu, D. R. *J. Am. Chem. Soc.* **2011**, *133*, 9972.
(26) Erben, A.; Grossmann, T. N.; Seitz, O. *Angew. Chem. Int. Ed.* **2011**, *50*, 2828.
(27) Erben, A.; Grossmann, T. N.; Seitz, O. *Bioorg. Med. Chem. Lett.* **2011**, *21*, 4993.
(28) Gillam, I. C.; Tener, G. M. *Anal. Biochem.* **1986**, *157*, 199.
(29) Viscidi, R. P.; Connelly, C. J.; Yolken, R. H. *J. Clin. Microbiol.* **1986**, *23*, 311.
(30) Keller, G. H.; Cumming, C. U.; Huang, D.-P.; Manak, M. M.; Ting, R. *Anal. Biochem.* **1988**, *170*, 441.
(31) Gillet, L. C. J.; Alzeer, J.; Schärer, O. D. *Nucleic Acids Res.* **2005**, *33*, 1961.
(32) Alzeer, J.; Schärer, O. D. *Nucleic Acids Res.* **2006**, *34*, 4458.
(33) Li, H.; Fedorova, O. S.; Trumble, W. R.; Fletcher, T. R.; Czuchajowski, L. *Bioconjugate Chem.* **1997**, *8*, 49.
(34) Lee, O. c.; Jeon, J.-H.; Sung, W. *Physical Review E* **2010**, *81*, 021906.
(35) Ambjörnsson, T.; Banik, S. K.; Krichevsky, O.; Metzler, R. *Biophys. J.* **2007**, *92*, 2674.
(36) Avignolo, C.; Roner, R.; Cai, S.; Bignone, F. A. *J. Biochem. Bioph. Methods* **1991**, *23*, 193.
(37) Draper, D. E.; Gold, L. *Biochemistry* **1980**, *19*, 1774.
(38) Kolb, H. C.; Finn, M. G.; Sharpless, K. B. *Angew. Chem. Int. Ed.* **2001**, *40*, 2004.
(39) Iha, R. K.; Wooley, K. L.; Nystrom̈, A. M.; Burke, D. J.; Kade, M. J.; Hawker, C. J. *Chem. Rev.* **2009**, *109*, 5620.
(40) Sletten, E. M.; Bertozzi, C. R. *Angew. Chem. Int. Ed.* **2009**, *48*, 6974.
(41) Liu, D. S.; Tangpeerachaikul, A.; Selvaraj, R.; Taylor, M. T.; Fox, J. M.; Ting, A. Y. *J. Am. Chem. Soc.* **2011**, *134*, 792.
(42) Liu, F.; Soh Yan Ni, A.; Lim, Y.; Mohanram, H.; Bhattacharjya, S.; Xing, B. *Bioconjugate Chem.* **2012**, *23*, 1639.
(43) Roloff, A.; Seitz, O. *Chemical Science* **2013**, *4*, 432.
(44) Rashidian, M.; Mahmoodi, M. M.; Shah, R.; Dozier, J. K.; Wagner, C. R.; Distefano, M. D. *Bioconjugate Chem.* **2013**, *24*, 333.
(45) Tijssen, P.; Kurstak, E. *Anal. Biochem.* **1984**, *136*, 451.
(46) Naganathan, S.; Ye, S.; Sakmar, T. P.; Huber, T. *Biochemistry* **2013**, *52*, 1028.
(47) Kodadek, T.; Duroux-Richard, I.; Bonnafous, J.-C. *Trends Pharmacol. Sci.* **2005**, *26*, 210.
(48) Claessen, J. H. L.; Witte, M. D.; Yoder, N. C.; Zhu, A. Y.; Spooner, E.; Ploegh, H. L. *ChemBioChem* **2013**, *14*, 343.
(49) Guillier, F.; Orain, D.; Bradley, M. *Chem. Rev.* **2000**, *100*, 2091.
(50) Gibson, S. E.; Hales, N. J.; Peplow, M. A. *Tetrahedron Lett.* **1999**, *40*, 1417.
(51) Telser, J.; Cruickshank, K. A.; Morrison, L. E.; Netzel, T. L. *J. Am. Chem. Soc.* **1989**, *111*, 6966.
(52) Böge, N.; Gräsl, S.; Meier, C. *J. Org. Chem.* **2006**, *71*, 9728.
(53) Swenson, M. C.; Paranawithana, S. R.; Miller, P. S.; Kielkopf, C. L. *Biochemistry* **2007**, *46*, 4545.
(54) Friedman, J. I.; Jiang, Y. L.; Miller, P. S.; Stivers, J. T. *Biochemistry* **2010**, *50*, 882.
(55) Sklyadneva, V. B.; Shie, M.; Tikchonenko, T. I. *FEBS Lett.* **1979**, *107*, 129.
(56) Huang, C. Y.; Miller, P. S. *J. Am. Chem. Soc.* **1993**, *115*, 10456.
(57) Semenyuk, A.; Darian, E.; Liu, J.; Majumdar, A.; Cuenoud, B.; Miller, P. S.; MacKerell, A. D.; Seidman, M. M. *Biochemistry* **2010**, *49*, 7867.
(58) Rajski, S. R.; Williams, R. M. *Chem. Rev.* **1998**, *98*, 2723.

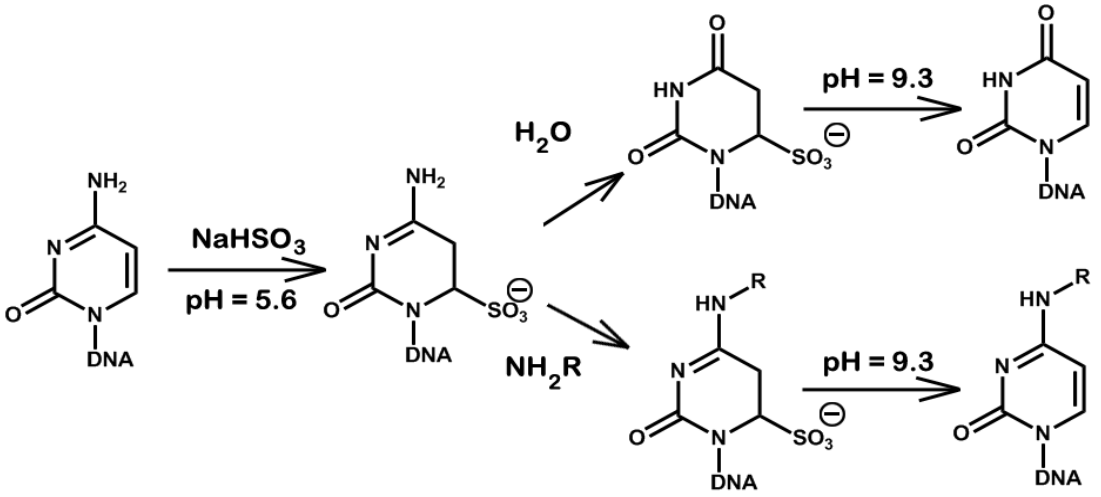
(59) Zhu, G.; Lippard, S. J. *Biochemistry* **2009**, *48*, 4916.
(60) Dieter-Wurm, I.; Sabat, M.; Lippert, B. *J. Am. Chem. Soc.* **1992**, *114*, 357.
(61) Gates, K. S.; Nooner, T.; Dutta, S. *Chem. Res. Toxicol.* **2004**, *17*, 839.
(62) Lindahl, T.; Nyberg, B. *Biochemistry* **1972**, *11*, 3610.
(63) Chetsanga, C. J.; Bearie, B.; Makaroff, C. *Chem. Biol. Interact.* **1982**, *41*, 217.
(64) Talman, E. G.; Brüning, W.; Reedijk, J.; Spek, A. L.; Veldman, N. *Inorg. Chem.* **1997**, *36*, 854.
(65) Shabarova, Z. A.; Bogdanov, A. A.; 1 edition (July 13, 1994) ed.; Wiley-VCH: 1994, p 589.
(66) Tanchuk, Y. V.; Gun'ko, V. M.; Roev, L. M.; Kornienko, A. A. *Theor. Exp. Chem.* **1981**, *16*, 434.
(67) Kočalka, P.; El-Sagheer, A. H.; Brown, T. *ChemBioChem* **2008**, *9*, 1280.
(68) Onizuka, K.; Taniguchi, Y.; Sasaki, S. *Bioconjugate Chem.* **2010**, *21*, 1508.
(69) Verma, S.; Eckstein, F. *Annu. Rev. Biochem* **1998**, *67*, 99.

CHAPTER 3

Hydrogensulfite-catalysed transamination of cytosine at the C4 position

3.1. Introduction

Hydrogensulfite-catalyzed transamination of cytosine is a well-known reaction which has been used to modify and probe the structure of nucleic acids[1]. The transformation of cytidine with hydrogensulfite has been described in detail[2,3]. Briefly, the reaction is most rapid at pH values between 4 and 6[4], and shows selectivity for C residues in single-stranded regions of nucleic acids. In the first step the fast hydrogensulfite addition to position C^6 of cytidine forms a reactive 6-sulfocytidine intermediate (**Scheme 3.1**). Nucleophiles, such as primary amines or hydrazine-derivative, attack the activated cytidine core and replace the N4-amino group. After incorporation of the new amino-derivative, the sulfo-group is hydrolyzed under alkaline conditions to restore the initial aromatic ring. The single-strand specificity stems from the fact that the site of hydrogensulfite-catalyzed nucleophilic replacement, i.e. the amino-group at position N4 in cytidine, is sterically hindered[4].



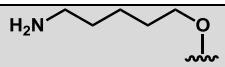
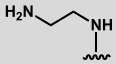
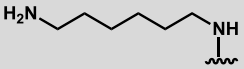
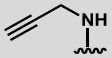
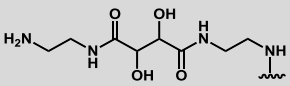
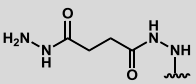
Scheme 3.1. Chemical transformation of cytosine is catalysed by hydrogensulfite as reported[5].

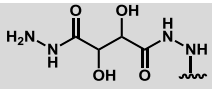
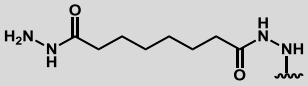
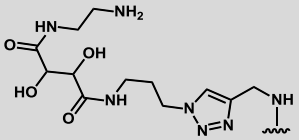
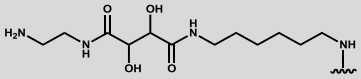
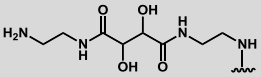
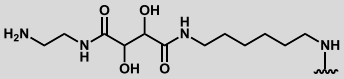
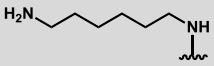
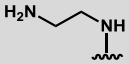
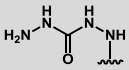
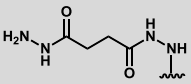
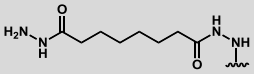
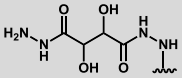
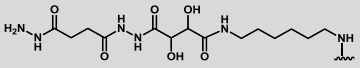
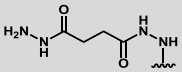
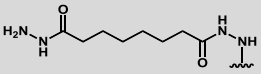
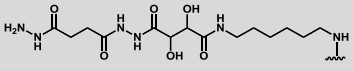
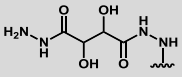
Till nowadays, all applications with hydrogensulfite catalyzed transamination of cytidine in nucleic acid chemistry were focused on the non-specific labeling of a DNA strand. Avignolo at al[6] treated a DNA single strand with ethylenediamine under transamination conditions to incorporate new functional amino groups in non-specific positions. In the next step they coupled the new amino-group with a NHS-biotin molecule and demonstrated effective DNA strand immobilization on a surface. Draper and co-workers[7] followed a similar strategy, but on the last step they used a fluorophore (with absorption maxima at λ = 360 nm and emission at λ = 465 nm) instead of biotin for fluorescent detection of the DNA concentration.

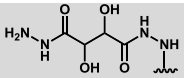
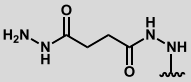
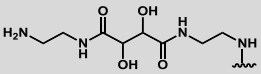
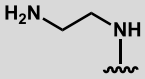
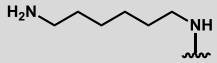
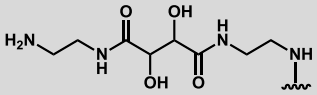
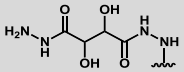
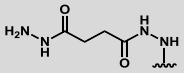
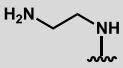
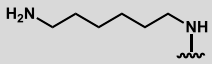
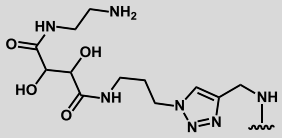
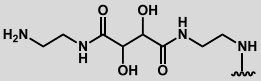
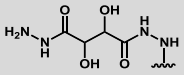
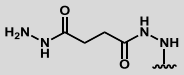
With our approach we demonstrate that site selective transamination of a cytosine base in an oligonucleotide can be achieved by using the modular system presented here. This system contains a donor sequence, which recognizes a specific region of a target single strand and forms a duplex structure. This places the reactive group, which is attached to the donor sequence *via* a linker arm, close in space to the target cytosine in the single stranded overhang of the target sequence. Prior to transamination the target cytosine needs to be activated with hydrogensulfite. Subsequently, nucleophilic replacement at the C4-position of cytosine leads to a cross-linked duplex, which can be chemoselectively cleaved with sodium periodate when a diol-containing linker is used to bridge the donor and target sequences. The newly formed aldehyde groups can be conjugated with molecules of interest, such as fluorophores or tags. In order to fully explore the sequence selectivity of the transamination cross-link methodology in terms of influence of complementary and neighboring bases, we tested different DNA sequences and demonstrated that the final results did not depend on the nucleotide composition of the complementary strand.

Sequences used:

*Complementary regions are underlined, (target C bases are indicated in bold, SRS: reactive sequence; ODN: template; ─C─: C located in the center of SRS).

Name	Reactive group	Position of ligation	sequence
SRS_3.1		PO_4	5’-CCC TTA ACC C-3’
ODN_3.1	─	─	5’-GGG AAT TGG G**C**T T-3’
ODN_3.2	─	─	5’-GGG AAT TGG GT**C** T-3’
ODN_3.3	─	─	5’-GGG AAT TGG GTT **C**-3’
SRS_3.2	─	PO_4	5’-CAG AGA AGG G-3’
SRS_3.3		C	5’-CAG AGA AGG G-3’
SRS_3.4		C	5’-CAG AGA AGG G-3’
SRS_3.5		C	5’-CAG AGA AGG G-3’
SRS_3.6		C	5’-CAG AGA AGG G-3’
SRS_3.7		C	5’-CAG AGA AGG G-3’

SRS_3.8		C	5’-CAG AGA AGG G-3’
SRS_3.9		C	5’-CAG AGA AGG G-3’
SRS_3.10		C	5’-CAG AGA AGG G-3’
SRS_3.11		PO_4	5’-CAG AGA AGG G-3’
SRS_3.12		PO_4	5’-CAG AGA AGG G-3’
SRS_3.13		C	5’-CAG AGA AGG G-3’
SRS_3.14		PO_4	5’-CAG AGA AGG G-3’
SRS_3.15		PO_4	5’-CAG AGA AGG G-3’
SRS_3.16		C	5’-CAG AGA AGG G-3’
SRS_3.17		C	5’-CAG AGA AGG G-3’
SRS_3.18		C	5’-CAG AGA AGG G-3’
SRS_3.19		C	5’-CAG AGA AGG G-3’
SRS_3.20		C	5’-CAG AGA AGG G-3’
SRS_3.21		PO_4	5’-CAG AGA AGG G-3’
SRS_3.22		PO_4	5’-CAG AGA AGG G-3’
SRS_3.23		PO_4	5’-CAG AGA AGG G-3’
SRS_3.24		PO_4	5’-CAG AGA AGG G-3’
ODN_3.4	–	–	5’-CCC TTC TCT G**C**T T-3’
ODN_3.5	–	–	5’-CCC TTC TCT GT**C** T-3’
ODN_3.6	–	–	5’-CCC TTC TCT GTT **C**-3’
SRS_3.25	–	PO_4	5’-CAA GTA GAA GAA GGG -3’

SRS_3.26		C	5’-CAA GTA GAA GAA GGG -3’
SRS_3.27		C	5’-CAA GTA GAA GAA GGG -3’
SRS_3.28		C	5’-CAA GTA GAA GAA GGG -3’
SRS_3.29		C	5’-CAA GTA GAA GAA GGG -3’
SRS_3.30		C	5’-CAA GTA GAA GAA GGG -3’
SRS_3.31		C	5’-CAA GTA GAA GAA GGG -3’
SRS_3.32		PO_4	5’-CAA GTA GAA GAA GGG -3’
SRS_3.33		PO_4	5’-CAA GTA GAA GAA GGG -3’
ODN_3.7	–	–	5’- CCC TTC TTC TCA TTG **C**AA-3’
SRS_3.34	–	–C–	GGT ATG GAT TAA CAA TTA GTG TGG T -3’
SRS_3.35		–C–	GGT ATG GAT TAA CAA TTA GTG TGG T -3’
SRS_3.36		–C–	GGT ATG GAT TAA CAA TTA GTG TGG T -3’
SRS_3.37		–C–	GGT ATG GAT TAA CAA TTA GTG TGG T -3’
SRS_3.38		–C–	GGT ATG GAT TAA CAA TTA GTG TGG T -3’
SRS_3.39		–C–	GGT ATG GAT TAA CAA TTA GTG TGG T -3’
SRS_3.40		–C–	GGT ATG GAT TAA CAA TTA GTG TGG T -3’
ODN_3.8	–	–	5’-ACC ACA CTA AAA A**C**A AAT CCA TAC C – 3’

3.2 Transamination of cytosine and analysis of reaction steps

In the beginning we tested the efficiency of transamination of cytosine with different nucleophiles, such as ethylenediamine, tris(2-aminoethyl)amine and glycine methyl ester. In the first step the fast hydrogensulfite additions to position C6 of cytidine forms a reactive 6-sulfocytidine intermediate (**Scheme 3.1**). To characterize and evaluate the reactivity of this intermediate, commercially available cytidine was incubated in presence of sodium hydrogensulfite at pH = 5.6 at RT. Reaction was monitored by RP-HPLC and after 6 hours almost quantitative transformation to 6-sulfocytidine was detected[8]. Final product was purified by RP-HPLC and the molecular mass measured with ESI-MS ($M+H^+_{found}$ = 326.1 Da, $M+H^-_{found}$ = 324.2 Da, M_{calc} = 325.06 Da) **(Fig. 3.1. C**).

Purified 6-sulfocytidine is quickly reverted to N^4-substituted cytidine in presence of amines or hydrolyzed to the deaminated nucleotide, i.e. thymidine, as a side reaction[9]. First we used tris(2-aminoethyl)amine (**Fig. 3.1. A**) for transamination. The choice of this model compound was stipulated in its multifunctionality and it is theoretically possible to obtain also cross-linked cytidine moieties at transamination conditions. The corresponding reaction results in the disappearance of the peak of 6-sulfonated cytidine at 11 min in the RP-HPLC chromatogram and the formation of a new broad peak with lower retention time at 10.2 min. ESI-MS measurements show two types of molecular ions responsible for deaminated ($[M+H]^+$ = 244.0 Da) and transaminated cytidines ($[M+H]^+$ = 373.2 Da) and hence no cross-linked cytidine moieties (**Fig. 3.1. C**). Deamination of 6-sulfo-cytidine is an important side reaction that results in formation of a carbonyl group at position C4.

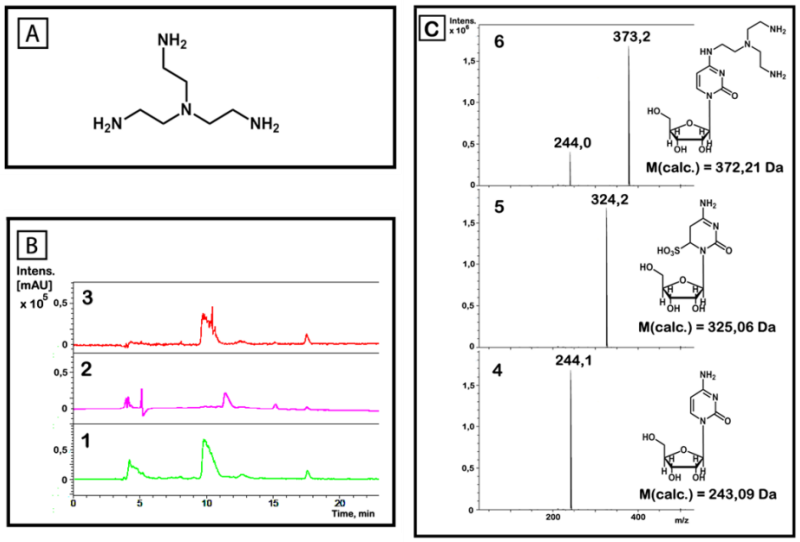


Figure 3.1. **A:** structure of tris(2-aminoethyl)amine. **B**: Reverse-phase HPLC chromatogram (λ = 260 nm) of **1:** cytidine, retention time 10.2 min, **2:** 6-amino-3-(4-hydroxy-5-hydroxymethyl-tetrahydro-furan-2-yl)-2-oxo-2,3,4,5-tetrahydro-pyrimidine-4-sulfonic acid (6-

sulfo-cytidine), retention time 11 min; **3:** 4-(([bis-(2-amino-ethyl)-amino]-methyl)-amino)-1-(4-hydroxy-5-hydroxymethyl-tetrahydro-furan-2-yl)-1H-pyrimidin-2-one (tris(2-aminoethyl)-cytidine), retention time 10.2 min. For the RP-HPLC conditions see Experimental part, "HPLC method 1". **C**: Deconvoluted ESI-MS spectra of **4**: cytidine (calc. mass 243.09 Da), **5**: 6-sulfo-cytidine, ESI-MS(-) (m/z): $[M-H]^-$ calcd. for $C_9H_{15}N_3O_8S$, 325.06 Da; found, 324.2 Da, **6**: tris(2-aminoethyl)cytidine, ESI-MS (m/z): $[M+H]^+$ calcd. for $C_{15}H_{28}N_6O_5$, 372.21; found, 373.2. All samples were measured in 0.01 M TEAA buffer at pH 5.6.

Next, we investigated the efficiency of nucleophilic replacement of the N4-amino-group in the 2-sulfo-cytidine intermediate. 0.2 mmol of 6-sulfo-cytidine was incubated with different concentration of tris(2-aminoethyl)amine, i.e. 0.1 eq., 0.2 eq., 0.5 eq., 1 eq., 2 eq., 5 eq. and 10 eq., in sodium acetate buffer at pH = 5.6. RP-HPLC analyses of each sample were done after 1 hour (**Fig. 3.2. A**) and 6 hours (**Fig. 3.2. B**). Efficiency of transamination was evaluated based on peak volume of 6-sulfo-cytydine and tris(2-aminoethyl)cytidine. Interpretation of these results demonstrated, that after 1 hour of incubation quantitate transamination require a large excess of amine nucleophile. In contrast, prolongation the reaction time to 6 hours demonstrated high efficiency of transformation with only 2-fold excess of nucleophile.

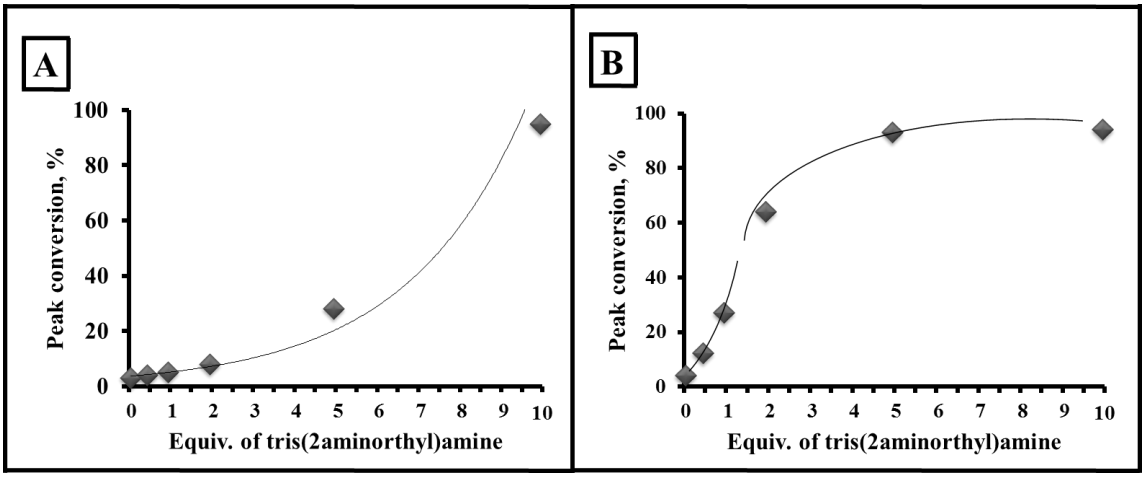
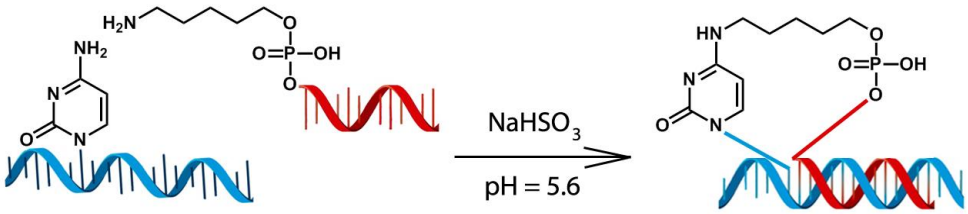


Figure 3.2. Formation of 6-tris(2-aminoethyl)cytidine in dependency of tris(2-aminoethyl)amine concentration. In the graphs percentage of 6-sulfo-cytidine conversion is plotted against tris(2-aminoethyl)amine concentration for two different incubation time **A:** 1 hours; **B:** 6 hours.

3.3 First generation modular system

Our previous results demonstrated that the efficient transamination of mononucleotides requires the presence of hydrogensulfite as catalyst as well as high concentrations of amino-derivatives in the last step. We hypothesized that attaching an amino-derivative covalently to the recognition sequence and guiding it to the target cytosine by formation of a rigid double helix structure will increase the effective concentration of the nucleophile in the direct vicinity of the target nucleotide to the level of concentrations which allow effective cross-link duplex formation.

Our first generation modular system for transamination involves conjugation of a primary amino group to the recognition sequence through a phosphordiester bond *via* a short flexible hexane linker. A ten nucleotide long DNA sequence was used as donor sequence (SRS_3.1), which allowed to form stable duplexes at room temperature ($T_{melting\ point}$ = 37 °C) and hence to recognize specific sequences in target DNA strands (**Scheme 3.2**).



Scheme 3.2. Interstrand cross-link formation between the first generation modular system and target templates. Donor sequence: SRS_3.1: 5'NH_2-CCC TTA ACC C-3', Template sequences: 5'-GGG AAT TGG G**XX X**-3' (target cytosine is placed instead of one "X").

Cross-linking experiments were carried out by mixing equimolar amounts of the ten nucleotide long reactive sequence (SRS_3.1) with three complementary 13-base template sequences varying in the position of cytidine in the overhang region (ODN_3.1, ODN_3.2, ODN_3.3). After incubation for 3 days in freshly prepared 1M sodium hydrogensulfite solution, reaction mixtures were analyzed with denaturing 18% polyacrylamide gel electrophoresis (PAGE) (**Fig. 3.3**). Only one DNA sequence (ODN_3.1) with cytidine close to the duplex region gave an additional single band with reduced electrophoretic mobility, consistent with the formation of a cross-linked duplex DNA. The yield in formation of cross-linked products was around 6%.

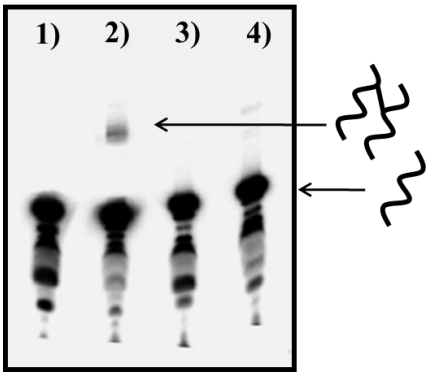


Figure 3.3. Cross-link formation between SRS_3.1 and 13-base templates under transamination conditions was analyzed by 18% denaturing PAGE: ODN_3.1 – lane 2, ODN_3.2 – lane 3, ODN_3.3 – lane 4). Lane 1 – shows ODN_3.1. in the absence of donor sequence.

We also applied size exclusion chromatography to purify the reaction mixture. In case of template ODN_3.1, the chromatogram recorded at λ = 254 nm demonstrated formation of an additional shoulder with higher retention time, than for initial unmodified templates (**Fig. 3.4**). Compounds eluting between 35 – 60 min were collected in five fractions and each fraction was characterized in two ways. Reverse-phase HPLC chromatograms demonstrated the presence of two peaks which correlate with SRS_3.1 (10.2 min) and unmodified ODN_3.1 (11.3 min) in the first fractions (**Fig. 3.5, samples 2** and **3**). These peaks decreased in intensities and disappeared in the next two fractions. In the last fraction that was correlated with newly formed shoulder RP-HPLC analysis demonstrated formation of a new peak at 10.8 min (**Fig. 3.5, sample 6**). In addition to the characterization of remaining SRS_3.1, MALDI-TOF MS analysis confirmed the presence of a cross-linked product with a modular mass of 7100.8 Da ($M_{calculated}$ = 7104.6 Da) (**Fig. 3.6**).

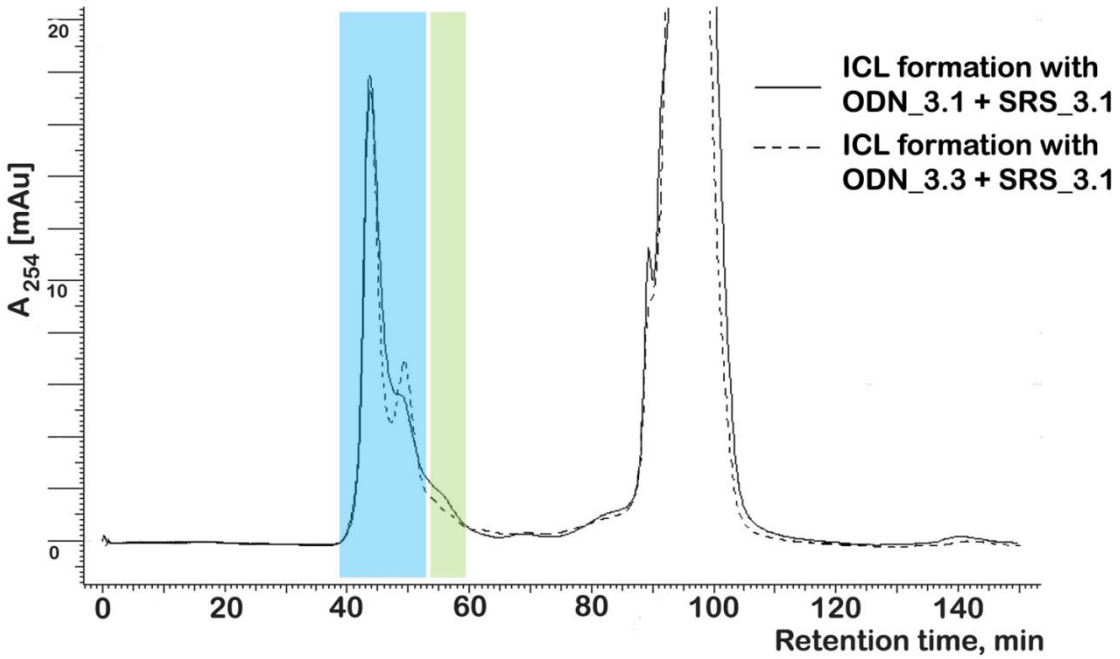


Figure 3.4. Elution profiles of size exclusion runs with Hiload 16/60 Superdex 75 prep grade column of templates with: ODN_3.1 - C in position n+1 (solid line) and ODN_3.3 - C in position n+3 (dashed line) after incubation with donor DNA. Peak at 45 min originates from free 13-mer template DNA, peak at 49 min from SRS_3.1 (both peaks indicated in blue square, fraction from 1 to 4 in the text), and shoulder at 56 min from cross-linked product (indicated in green square, fifth fraction). Retention times of donor and target oligonucleotides were identified by separate injection of single compounds.

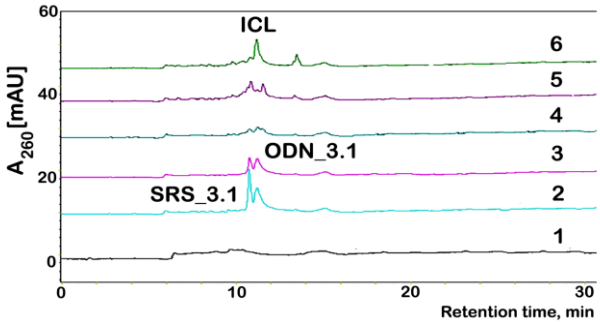


Figure 3.5. Reverse-phase HPLC chromatograms (λ = 260 nm) of fractions after size exclusion purification (**Fig. 3.4**). 1 – fraction at 35-40 min, 2 – 40-45 min, 3 – 45-47 min, 4 – 47-52min, 5 – 52-55 min, 6 – 55-60 min. Peak at 10.2 min originates from unreacted SRS_3.1, 11.3 min – unreacted ODN_3.1, 10.8 min – cross-linked strands. Retention times of SRS_3.1 and ODN_3.1 were identified by separate injection of single compounds.

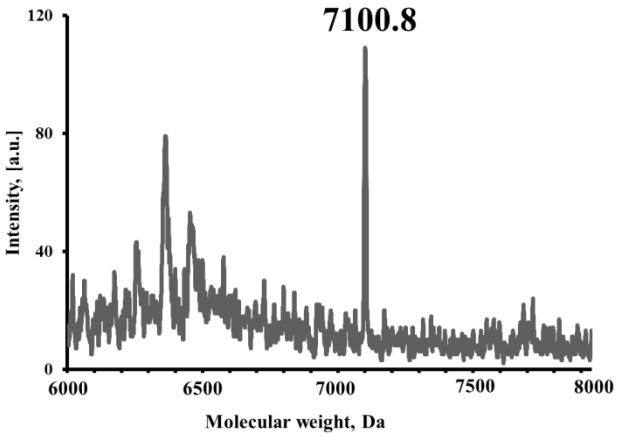


Figure 3.6. MALDI-TOF spectrum of cross-linked product after RP-HPLC purification of the 5-th fraction obtained from size-exclusion chromatography.

After these first promising results we designed series of new reactive groups for the transamination reaction and also evaluated reactive sequences in which the reactive group and linker is attached *via* a transamination reaction to the 5'-cytidine base of the recognition sequence instead of linking it to the phosphate. We also checked a number of different bifunctional amines[10] as reactive groups. According to our previous results, quantitative transformation of the NH_2 group at the C4 position to different modified amines gave us a tool for easy modification of a donor strand with a high variety of molecules. Yields of modified donor molecules were in the range of 60– 95% (**Table 3.1**).

First we focused on reducing the steric hindrance and used ethylenediamine as a short non-cleavable linker (SRS_3.3). The donor sequence was incubated according to the original protocol with three appropriate templates (ODN_3.4, ODN_3.5, ODN_3.6). After an incubation period of 3 days, gel electrophoresis was performed. Similar to previous results, only the template with cytidine in position next to the duplex region (ODN_3.4) provided efficient cross-link formation with a yield of 11% (**Fig. 3.7**). The two other templates only gave negligible amounts of product with yields of 2% and 5%, respectively. For a more precise analysis of this reaction, we repeated the reaction at higher concentration using the same protocol and purified it by PAGE. The slower running band was analyzed by MALDI-TOF (**Fig. 3.8**) and the observed molecular weight of 7003.7 Da was consistent with the calculated molecular weight of the cross-linked oligonucleotide, i.e. 7001.8 Da. The lower band contains unmodified template (ODN_3.4) as well as reactive sequence (SRS_3.3). Our attempts to purify reactions by RP-HPLC or size-exclusion chromatography were unsuccessful due to overlapping peaks of cross-linked product with initial compounds.

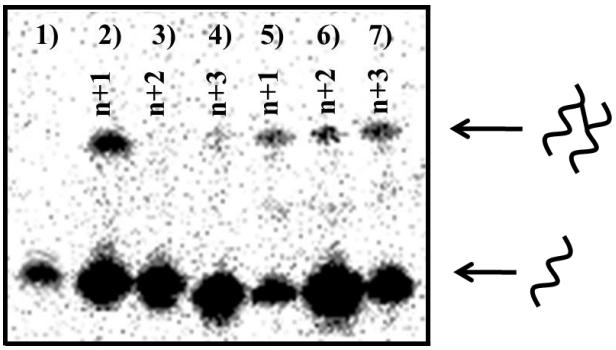


Figure 3.7. Phosphorimage autoradiogram of 18% denaturing PAGE analysis of hydrogensulfite catalyzed cytidine transamination using a sequence containing a ethylenediamine linker (SRS_3.3) and a complementary template with C in posinion n+1 (ODN_3.4, lane 2), n+2 (ODN_3.5, lane 3) and n+3 (ODN_3.6, lane 4). The analogues reaction using a reactive sequencewith a hexylenediamine linkers (SRS_3.4) was carried out with the same templates (lane 5-7). Incubation ODN_3.4 in absence of reactive sequence is shown in lane 1.

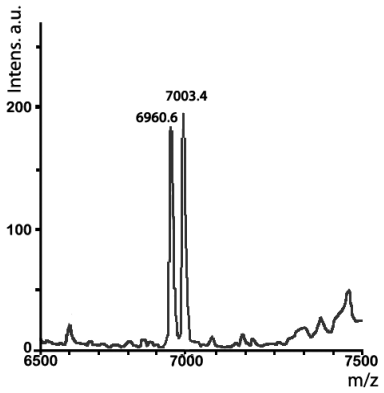


Figure 3.8. MALDI-TOF spectrum showing the cross-linked DNA duplex formed between donor molecule with ethylenediamine linker (SRS_3.3) and template with cytosine at position n+1 in overhang region (ODN_3.4) ($M_{observe}$ = 7003.4 Da, M_{calc} = 7001.8 Da). Ion peak at 6960.6 Da is responsible for degradation in template strand.

In the next step a cleavable linker was simulated by increasing the linker length and flexibility while at the same time preserving a similar electronic structure with respect to the ethylenediamine linker. 1,6-diaminohexane fulfills this requirement and was introduced to the 5'-end cytidine as a reactive group (SRS_3.4). Site-specificity in these reactions was lower than with the shorter linker, as the cross-linking band was consistently observed for all three duplexes (**Fig. 3.7**). The highest yield was observed with cytosine at position n+1 that equals to 7%. For ODN_3.5 efficiency was 3% and for ODN_3.6 6%. The slight increase in yield with the last template can be explained as following. The target cytidine occupies the last position in the oligonucleotide sequence, resulting in decrease it steric hindrance and an eases hydrogensulfite activation as well as attaching of the amino-reactive group from SRS_3.4.

To be able to quantify reaction yields we added an unmodified oligonucleotide as internal standard to the mixture after reaction. RP–HPLC analysis showed that the reactive oligonucleotide strand (SRS_3.4) at 27.6 min was almost consumed after 20 days (**Fig. 3.9. A**), but no clear peak corresponding to a cross-linked product could be observed in contrast to the results from PAGE. For correct evaluation of peak distribution, we integrated peaks corresponded to SRS_3.4 and ODN_3.4 after each experiment. Normalization of the two peaks were done based on the internal standard. After integration and normalization we observed that the peak at 29.1 min assigned to unmodified ODN_3.4 increased in intensity (**Fig. 3.9. B**). As additional argument, this peak was purified by RP-HPLC and identified by MALDI-TOF MS. In spite of the poor ionization efficiencies of the cross-linked intermediate MALDI-TOF demonstrated formation of the cross-linked product ($M_{calc.}$ = 7059.4 Da) and the presence of unmodified ODN_3.4 (M = 3829.9 Da) (**Fig. 3.9. D**). RP-HPLC analysis of all cross-linking reactions (with ODN_3.5 and ODN_3.6) demonstrated similar behavior in peaks transformation (**Fig. 3.9. C**) that allows us to propose, that cross-linked products have similar retention times as the unmodified templates.

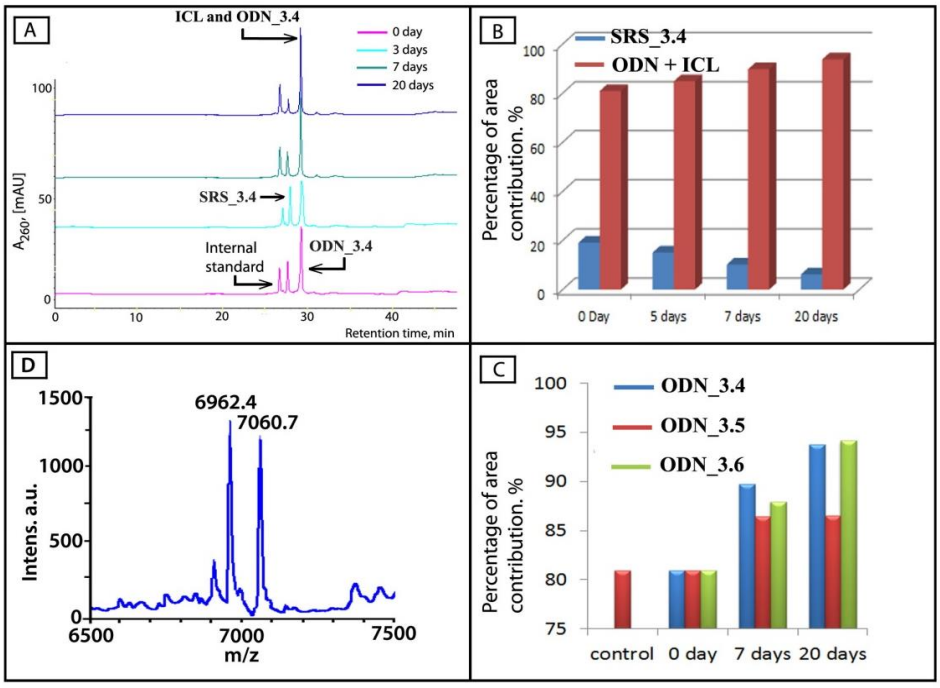


Figure 3.9. Cross-linking of hexylendiamino-modified donor DNA (SRS_3.4) with a template which contains target cytosine at position n+1 (ODN_3.4), **A**: Reverse-phase HPLC chromatograms of aliquots from reaction mixture were taken after 0, 3, 7 and 20 days. **B**: Integration of two peaks and their normalization to internal inert standard provide average crosslinking efficiencies, based on two experiments. **C**: Comparative graph of cross-linked product formation between donor molecule and three DNA templates. **D:** MALDI-TOF spectrum represented a molecular ion of cross-linked DNA duplex between donor molecule with

hexylendiamine linker and template with cytosine at position n+1 in overhang region ($M_{observe}$ = 7060.4 Da, M_{calc} = 7059.4 Da). Ion peak at 6962.4 Da is responsible for degradation in template strand.

Thermal stability of the different cross-linked duplexes was further analyzed by UV-melting experiments. Previous analysis showed that a cross-linked duplex is quite strongly stabilized compared to a non-modified duplex. First, the peak with a retention time of 29.1 min after RP-HPLC purification, which contains both initial template (ODN_3.4) and cross-linked duplex, was evaluated (**Fig. 3.9 A**). The melting curve reveals a single transition point at 46 °C (**Fig. 3.10, green curve**) that is more stable than the non-cross-linked duplex (blue curve, T_m = 39 °C). As was shown in MALDI-TOF experiment (**Fig. 3.9**), HPLC purified peak contains a mixture of cross-linked product and unmodified ODN_3.4. The presence of unmodified ODN_3.4 next to the cross-linked duplex was verified by addition of SRS_3.2 to the mixture. The melting curve of this mixture indeed reveals two transition points, one at 47 °C attributed to the cross-linked duplex between SRS_3.4 and ODN_3.4, and a second one at 36 °C, originating from melting of the duplex between SRS_3.2 and unmodified ODN_3.4 (**Fig. 3.10, violet curve**). The ethylenediamine covalent bridge between two strands increases the duplex stability by 6°C compared to the hexylenediamine covalent bridge (red curve). The sample was purified with PAGE analogues to **Fig. 3.8** before measurement.

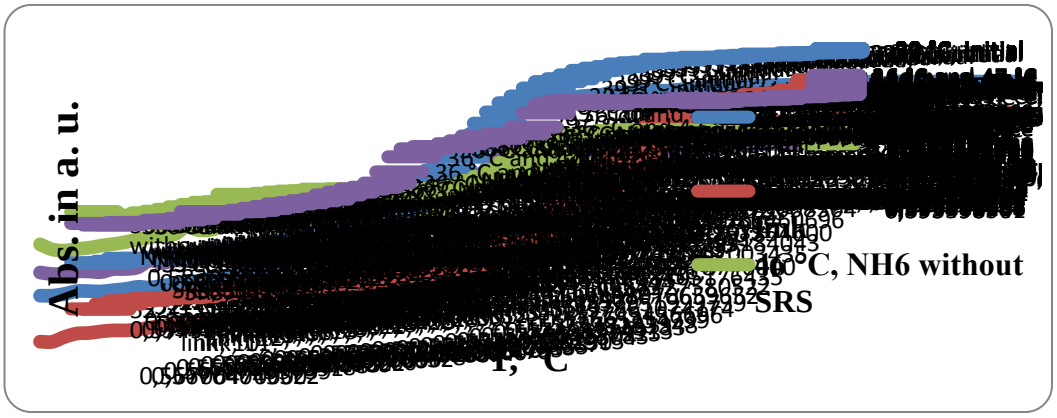


Figure 3.10. Comparison of melting curves for: blue line - non-modified duplex, i.e. SRS_3.3 and ODN_3.4 before cross-link formation; red line - cross-linked duplex between SRS_3.3 and ODN_3.4 (with ethylenediamine linker); green line - cross-linked duplex between SRS_3.4 and ODN_3.4 (with hexylenediamine linker). The sample also contained unmodified ODN_3.4 (retention time 29.1 min on RP-HPLC chromatogram (**Fig. 3.9**); violet line – same sample as for the green curve but with addition of SRS_3.2. Melting curves were recorded at λ = 260 nm with a heating rate of 1 °C/min. Duplex concentration was 2 μM in 100 mM NaCl and 10 mM acetate buffer at pH 7.

3.4 Stability of sulfonated DNA helix

Nucleophilic substitution at the C4-position of cytosine requires addition of a hydrogensulfite ion to the 5,6-double bond and leads to incorporation of an additional negative charge to the DNA strand and hence destabilizes the helix. The ability of template oligodeoxyribonucleotides and their sulfonated derivatives to form duplexes with the complementary donor DNA strand was studied. The 13 nucleotides long DNA single strand ODN_3.4, which contains six cytidines was incubated for 6 hours with 1 M $NaHSO_3$ at pH = 5.6 and then dialyzed against 5 mM sodium acetate pH = 5.6. The sulfonated DNA was isolated and its hybridization features were tested by UV spectrophotometric determination of melting curves (**Fig. 3.11**). The duplexes between sulfonated ODN_3.4 and SRS_3.2 in acetate buffer at pH = 5.6 did not show any transition (**Fig. 3.11, green curve**).

In the next step we remove the sulfone groups from cytosine core by hydrolysis under alkaline pH. As previously mentioned[11], the sulfonated group in cytosine shifts absorption maxima towards infrared wavelengths with new maxima at λ = 275 nm. Kinetic measurement showed recovering of initial absorption spectra at λ = 254 nm and reconstruction of the cytosine core by removing sulfonated group within two hours. Remeasuring UV melting plot demonstrated appearance of the transition point similar to unmodified samples (blue curve versus red curve). Due to these results we assumed that the addition of sulfonate groups destabilizes DNA duplex formation and decreases the yield of cross-linked product formation.

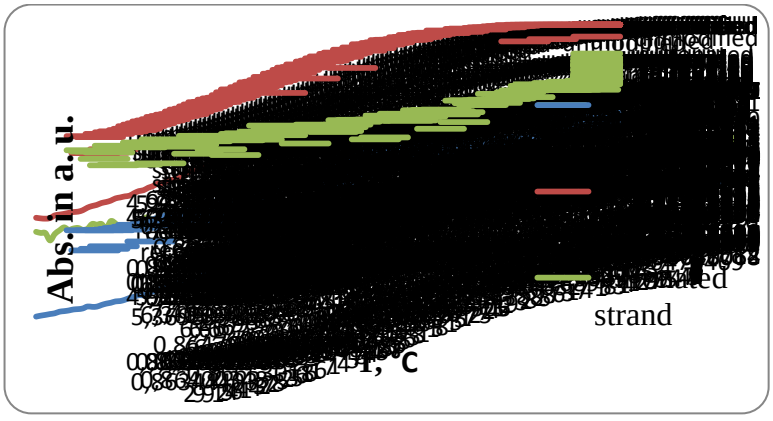
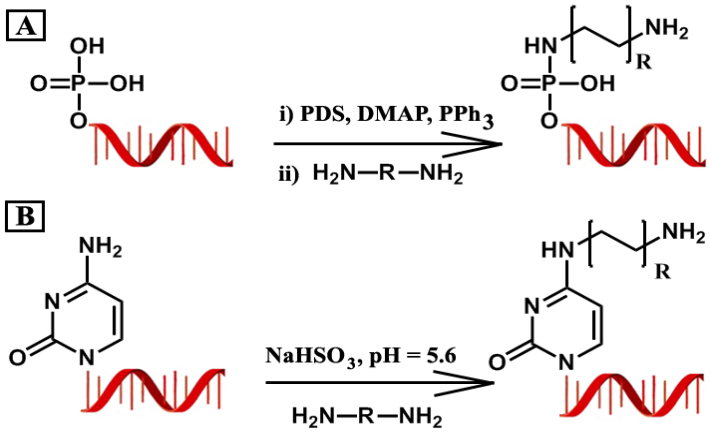


Figure 3.11. Comparison of melting curves for SRS_3.2 with complementary oligonucleotide ODN_3.4 (red curve, T_m = 24 °C), sulfonated ODN_3.4 or desulfonated (regenerated) ODN_3.4 (blue curve, T_m = 25 °C). Melting curves were recorded at λ = 260 nm with a heating rate of 0.5 °C/min. Duplex concentration was 2 μM in 100 mM NaCl and 10 mM acetate buffer at pH 5.6.

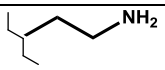
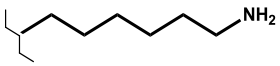
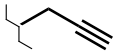
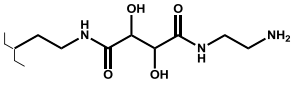
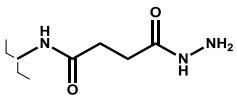
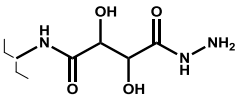
3.5 Design of cleavable linker

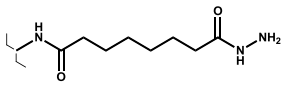
To allow a detailed evaluation of cross-linking selectivity, a series of linkers were incorporated in the donor sequence which can be separated into two different types (**Scheme 3.3**). In one, the linkers were attached to recognition sequences through phosphate-moiety and in the other through 5'-cytosine base (**Table 3.1**).



Scheme 3.3. Synthesis of donor molecule by two strategies: **A:** formation of phosphoramidite bond between amino-containing linker and recognition sequence[12]; **B:** transamination of the N4 position of a cytosine base at the 5'-end of the recognition sequence..

Table 3.1. N^4-amino-cytidine modified donor sequence synthesized from a single cytidine containing DNA primer and amino-derivatives by activation with sodium hydrogensulfite.

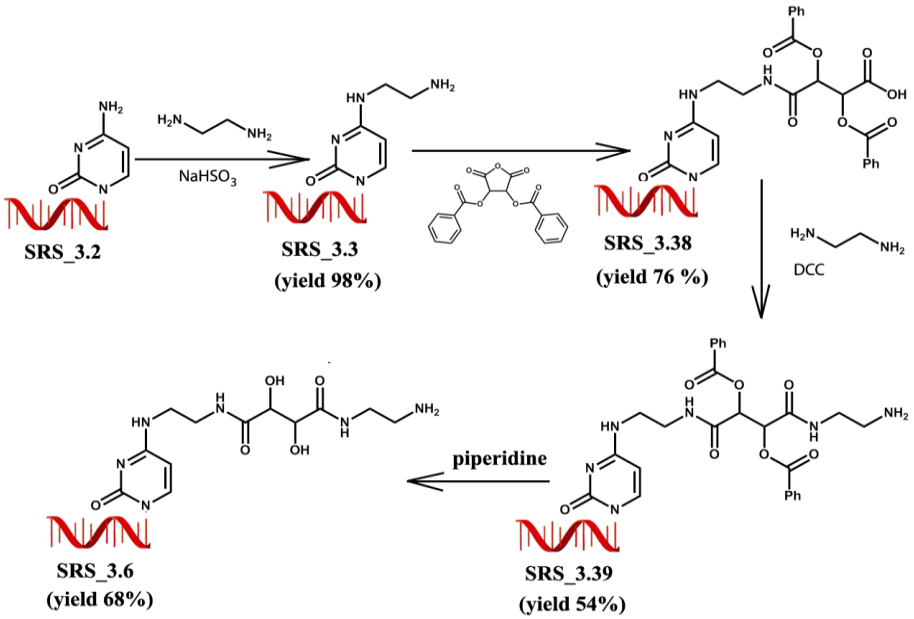
Entry	Name of reactive sequence	R	M_{calc}, Da	M_{found}, Da	Yield[1], %
1	SRS_3.3		3169.2	3168.4	98
2	SRS_3.4		3226.0	3223.3	75
3	SRS_3.5		3165.2	3163.1	95
4	SRS_3.6		3344.1	3341.9	60
5	SRS_3.7		3256.3	3258.0	92
6	SRS_3.8		3288.3	3291.8	65

7	SRS_3.9		3312.3	3314.2	70

1. Compounds SRS_3.5 and SRS_3.6 were purified by RP-HPLC and the final concentrations of the strands in the purified fractions were determined according to their absorption maxima at 254 nm. For other compounds approximated yield were based on the intensity of appropriate peaks during MALDI-TOF MS measurements.

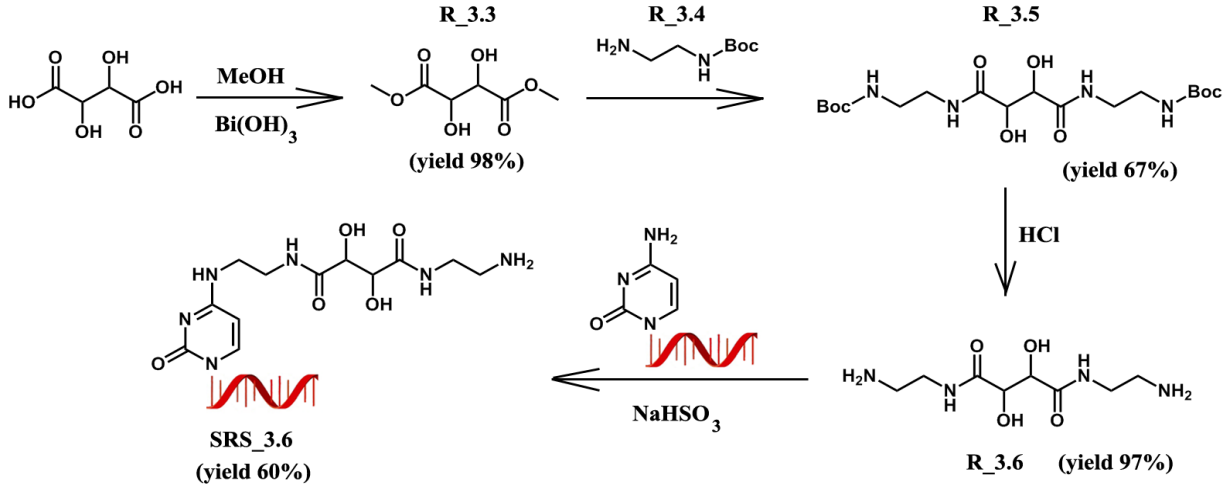
The traditional route of attaching different reactive groups through the sugar/phosphate moiety requires incorporation of a linker during solid phase synthesis. To make our system more simple, less expensive and independent from solid-phase modifications we chose an alternative route, which was proposed by Boutorine[12] (**Scheme 3.3 A**). The basic idea of this approach is the formation of a phosphoramidate bond between the 5'-phosphate group of a commercially available oligonucleotide and the amino-group of different reagents. The final product can be easily precipitated and purified by chromatography.

For the incorporation of reactive groups and linkers through nucleobases we used our previously gained experience with the transamination of cytosine (**Scheme 3.3 B**). Several approaches to introduce a diol-cleavable linker into DNA have been tried. In the beginning we tried to build up the linker piece by piece[13], starting from an unmodified DNA primer with a 5'-cytosine base **(Scheme 3.4**). In the first step the donor DNA primer (SRS_3.2) was transaminated with ethylenediamine as described earlier. Next, aminomodified DNA (SRS_3.3) was incubated in DMSO-H_2O mixture with diphenyltartaric anhydride (R_3.1) which led to incorporation of a carboxylic group for further modifications. Amide coupling by DCC (SRS_3.39) and deprotection of hydroxygroups with 10 eq. excess of piperidine in phosphate buffer at pH = 9.3 produced the desired reactive sequence (SRS_3.6). Despite many advantages this method demonstrated only traces of final product due to amine coupling under aqueous conditions and final deprotection steps. The overall yield was approximately 27%.



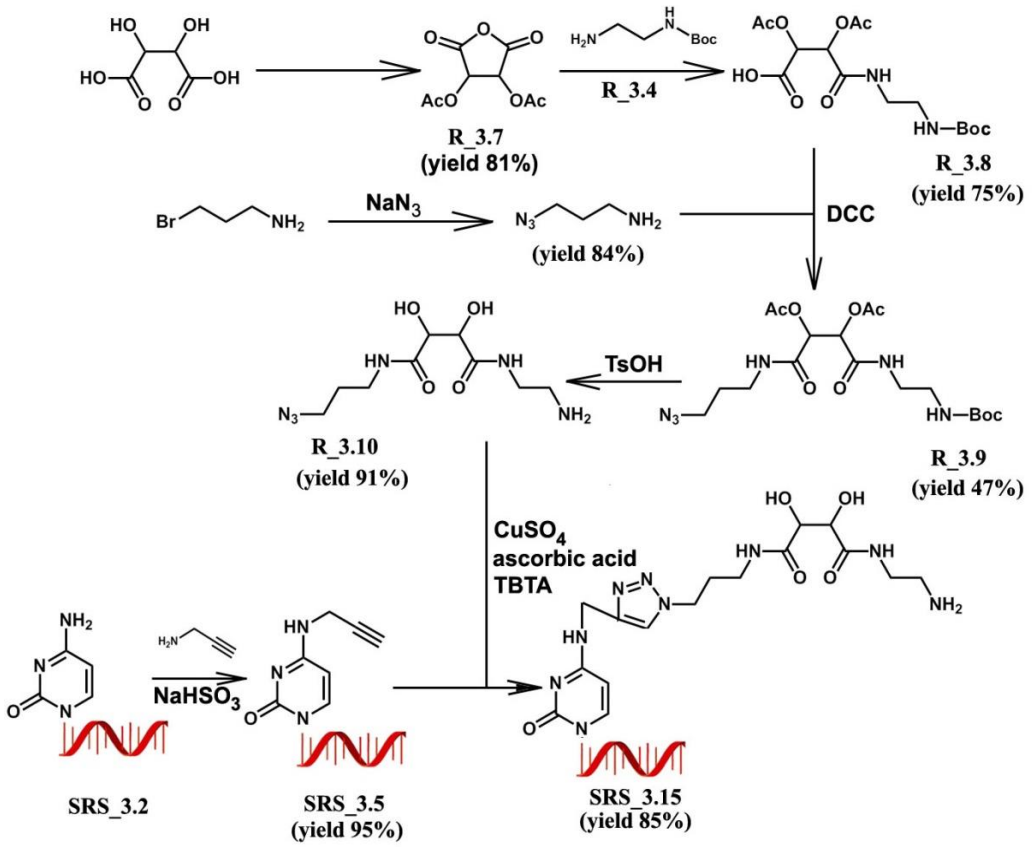
Scheme 3.4. "Step-by-step" synthesis of reactive sequence which contains symmetrical diol cleavage site in the linker. For detailed conditions see Experimental part and in the text.

An alternative synthetic route is based on the modular synthesis of a water soluble symmetrical diol intermediate[14] with two primary amino-groups and its subsequent incorporation into the recognition sequence without the need for deprotection steps (**Scheme 3.5**). In the first step, tartaric acid was converted to its dimethyl esters (R_3.3) by selective esterification catalysed by boric acid. In a typical reaction the carboxylic acid was dissolved in methanol, then boric acid was added, and the reaction was stirred at room temperature for 18 hours. In the next step, dimethyl-tartrate was refluxed with 2 eq. of mono-N-Boc-ethylenediamine (R_3.4) in MeOH, which provided the Boc-protected tartaric acid diamide R_3.5 in 67% yield. Treatment of the diamide with a solution of HCl in 1,4-dioxane to remove the Boc-groups, followed by purification of the resulting diamine hydrochloride salt, yielded pure monomer in quantitative yield (R_3.6). The last step of synthesis includes direct transamination of the 5'-end cytosine nucleobase in the recognition sequence (SRS_3.2) with R_3.6 that results in formation of reactive sequence (SRS_3.6) in 60% yield. Similar to transamination of mononucleotide, 10 eq. excess of diamino-containing linker prevents dimer formation *via* transamination of second recognition sequence (DNA – linker – DNA). No dimer was observed during MALDI-TOF MS experiments (see Experimental part).



Scheme 3.5. Synthesis of symmetrical diol-containing cleavable linker and its incorporation into recognition DNA sequence SRS_3.2 by hydrogensulfite catalysed transamination of cytosine. For detailed information see Experimantal part and text.

The third approach relies on the preparation of an unsymmetrical tartaric acid derivative with only a single primary amine and a bioorthogonal azide group at the other end for coupling with the donor sequence (**Scheme 3.6**). To prepare this compound, we used the well-known di-O-acetyl protected tartaric anhydride (R_3.7), which was prepared from L-tartaric acid in 81% yield by refluxing in acetic anhydride in the presence of sulfuric acid[15]. The anhydride was subsequently converted to the N-ethylamino-tartaramic acid (R_3.8) in 75% yield by treatment of the anhydride with a solution of Boc-protected ethylenediamine in THF. This sequence seemed more practical than the attempt of selective monoamidation of L-tartaric acid with non-protected diamines. Standard coupling conditions using DCC were used for amide bond formation between tartaramic acid and azido-aminopropane to provide the unsymmetrical tartrat in 47% yield (R_3.9). Simultaneous removal of the Boc- and acetyl- protecting groups was then achieved with 5 eq. of p-toluenesulfonic acid monohydrate in a CH_2Cl_2-MeOH solvent system. The crude product resulting from this reaction was purified and neutralized by ion exchange resin, giving the unsymmetrical tartaric acid in 26% of overall yield starting from initial tartaric acid (R_3.10). In the parallel synthesis we modified 5'-end cytosine containing recognition sequence (SRS_3.2) with 2-aminopropyne *via* hydrogensulfite catalysis. Coupling between unsymmetrical diol linker and alkyne-containing DNA primer (SRS_3.5) was performed in presence of Cu(I) that gives final compound (SRS_3.10) in 80% yield (yield was calculated according oligonucleotide sequence). Final product was purified by RP-HPLC and MALDI-TOF experiment demonstrated molecular ion of desired compound (M_{found} = 3436.3 Da, see Experimental part).

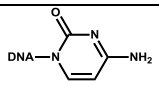
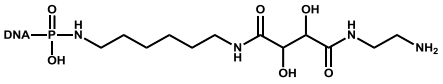
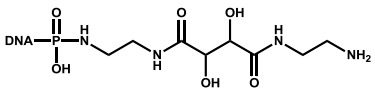
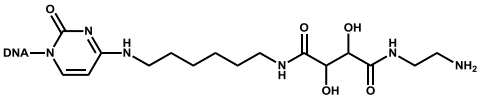
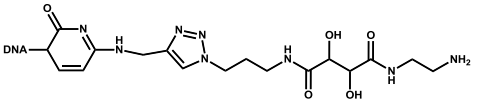
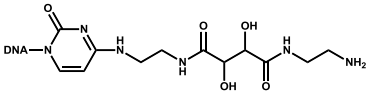
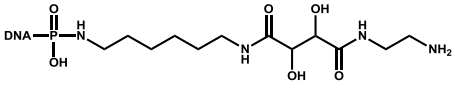


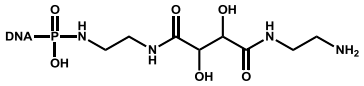
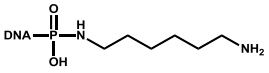
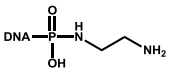
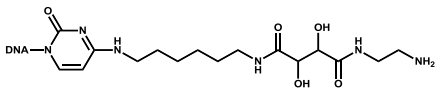
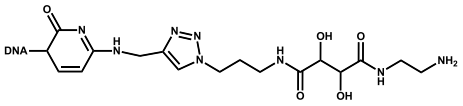
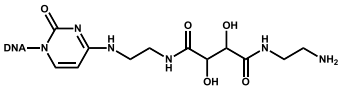
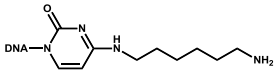
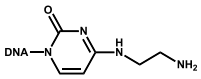
Scheme 3.6. Synthesis of unsymmetrical diol-containing cleavage module and its incorporation into recognition DNA sequence *via* [1+3] polar Huisgen coupling.

Next, we tested the influence of the linker length on localization reactive group in the space and efficiency in formation of cross-link product. To get this we designed a library of linkers that contain a diol cleavage site between the recognition sequence and the reactive group (**Table 3.2**). In addition, the spacer between diol and reactive group contains two carbon atoms in all compounds. This size should provide flexibility to reactive group in proximity to target cytosine core and allow effective nucleophilic replacement. At the same time, after cleavage a short side chain is introduced into the target DNA strand. We also varied the linker length between diol group and recognition sequence from 2 to 6 carbon atoms in the chain. Additionally, the reactive group with the linker was attached to the recognition sequence in two different ways. Firstly, the group can be attached in the usual way *via* an amino-group to either the phosphate moiety or the cytosine N4 position. Secondly, an azide group can be introduced into the linker as depicted in **Scheme 3.6** and applied for conjugation to an alkyne group in a "click-reaction" to any type of biomolecule. These three synthetic routes allow us to generate a small library of linkers. Compounds SRS_3.11 and SRS_3.13 were synthetized by "step-by-step" route, compounds SRS_3.3, SRS_3.4, SRS_3.6, SRS_3.12, SRS_3.14 and SRS_3.15 – by modular route with amino-binding group, and compound SRS_3.10 by modular route with azide-binding group.

Transamination reactions with these reactive sequences were performed as described above: reactive sequences were mixed with complementary DNA template (ODN_3.4) in which the target cytidine was placed in (n+1) position to reactive group and incubated for three days in the presence of sodium hydrogensulfite. However, cross-link formation was only observed with some compounds (**Fig. 3.12**). Compounds SRS_3.4 and SRS_3.6 with the ligand attached to cytosine demonstrated almost similar efficiencies in formation of covalent bound product (3% and 5% respectively). This resulted in completely ineffective reactions with linkers longer than 12 atoms (SRS_3.13, SRS_3.10). On the opposite side, clear cross-link formation was observed for samples SRS_3.11, SRS_3.12, SRS_3.14 and SRS_3.15, where linkers were incorporated to the phosphate moiety and no dependence of linker size on cross-link formation was detected. Additionally, diol-containing samples were oxidized by sodium periodate (lanes 2-6) to demonstrate the cleavage of the interstrand covalent bond. After oxidation, only DNA primers with electrophoretic mobility similar to the unmodified DNA template were detected in agreement with formation of modified single stranded DNA templates. According to the original protocol, the cleavage is most efficient at acidic pH (pH = 4.5) and lasted for 2 – 4 hours[16],[17]. In this series of experiment we incubated samples only for 1 hour at pH = 6.2 resulting in not quantitative cleavage of cross-link and hence traces of the upper band are still visible (1% - 2%).

Table 3.2. List of linker arms which were used in cross-link experiment.

№ on PAGE	№ of compound	R	Yield of synthesis	$NaIO_4$
1	SRS_3.2		—	
2	SRS_3.11		52	+
3	SRS_3.12		87	+
4	SRS_3.13		41	+
5	SRS_3.10		80	+
6	SRS_3.6		60	+
7	SRS_3.11		37	

8	SRS_3.12		86
9	SRS_3.14		quantitate
10	SRS_3.15		quantitate
11	SRS_3.13		41
12	SRS_3.10		80
13	SRS_3.6		60
14	SRS_3.4		quantitate
15	SRS_3.3		quantitate

№ of lane	1		2	3	4	5	6	7	8	9	10	11	12	13	14	15
№ of compound	SRS_3.2		SRS_3.11	SRS_3.12	SRS_3.13	SRS_3.10	SRS_3.6	SRS_3.11	SRS_3.12	SRS_3.14	SRS_3.15	SRS_3.13	SRS_3.10	SRS_3.6	SRS_3.4	SRS_3.3
$NaIO_4$			+	+	+	+	+									
Yield in ICL formation, %			0.3	1.5	1.1	0.4	0.8	1.6	2.5	4	5.2	0.4	0.3	5	3	9

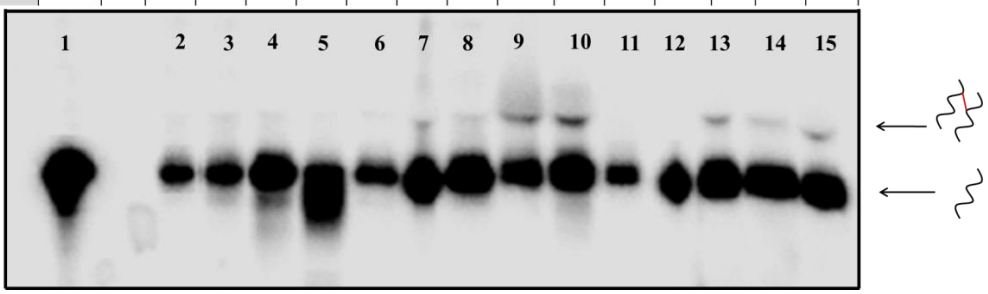
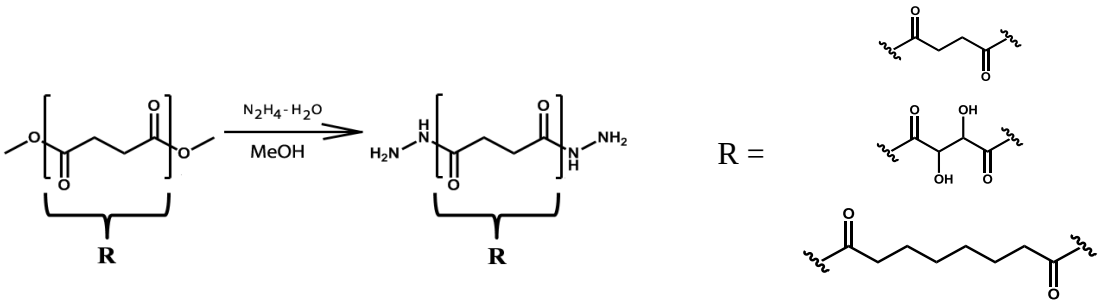


Figure 3.12. Reaction of 10-mer amino-containing reactive sequences with the complementary 13-mer target oligonucleotide ODN_3.4 was analyzed by 12% denaturing PAGE. Recognition sequences were modified with different linkers through phosphoramidate bond (lanes 7-10) or through the 5'-cytosine (lanes 11-15) and incubated with template under transamination conditions. After 3 days products with cleavable linkers were oxidized with sodium periodate (lanes 2-6) (see also **Table 3.2**).

3.6 Second generation modular system

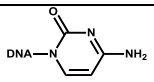
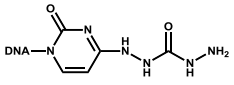
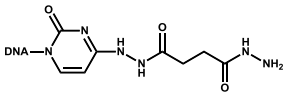
Data from several groups shows that also hydrazine, semicarbazide or acetylhydrazide can replace the amino group of cytidine[18,19]. The transamination reaction also occurs with cytosine residues in DNA and RNA single strands when catalyzed with hydrogensulfite. Based on this data we have developed a second generation molecular system with increased nucleophilic properties of the reactive group. Replacement of the amino by hydrazine group can be achieved by refluxing the symmetric dimethyl ester of the respective acid with hydrazine (**Scheme 3.7**).

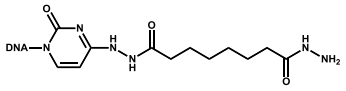
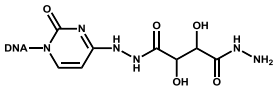
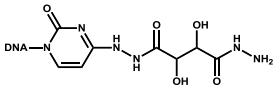
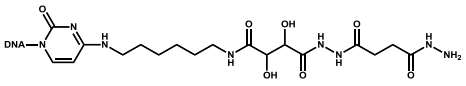
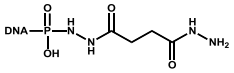
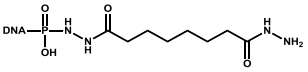
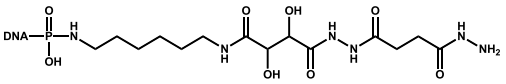
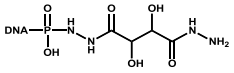
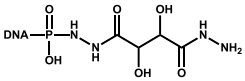
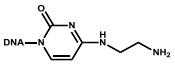


Scheme 3.7. Synthesis of hydrazine containing reactive groups. For details see Experimental part.

Hybridization experiments were repeated under the same conditions using the different hydrazine-modified reactive sequences depicted in **Table 3.3**. For hydrazine-building blocks, a significantly enhanced cross-linking reactivity can be observed based on 12% PAGE analysis (lanes 3-6 and 9-12) (**Fig. 3.13**). Similar to amino-containing reactive sequences, hydrazine derivatives with linkers longer than 12 atoms that were attached to the 5'-cytosine (SRS_3.20) did not show formation of ICL.

Table 3.3. List of linkers and reactive groups for cross-linking experiments and yields of their synthesis.

Entry	sequence		$NaIO_4$	Yield of SRS, %
1	SRS_3.2			–
2	ODN_3.4		NaHSO_3	–
3	SRS_3.16			55
4	SRS_3.17			92

5	SRS_3.18			65
6	SRS_3.19			89
7	SRS_3.19		+	–
8	SRS_3.20			37
9	SRS_3.21			79
10	SRS_3.22			68
11	SRS_3.23			45
12	SRS_3.24			67
13	SRS_3.24		+	–
14	SRS_3.3			98

№ of lane	1	2	3	4	5	6	7	8	9	10	11	12	13	14
№ of compound	SRS_3.2		SRS_3.16	SRS_3.17	SRS_3.18	SRS_3.19	SRS_3.19	SRS_3.20	SRS_3.21	SRS_3.22	SRS_3.23	SRS_3.24	SRS_3.24	SRS_3.3
$NaIO_4$							+						+	
Yield in ICL formation, %	0.7	0.3	21	23	19	18	0.4	0.5	19	21	17	9	0.7	9

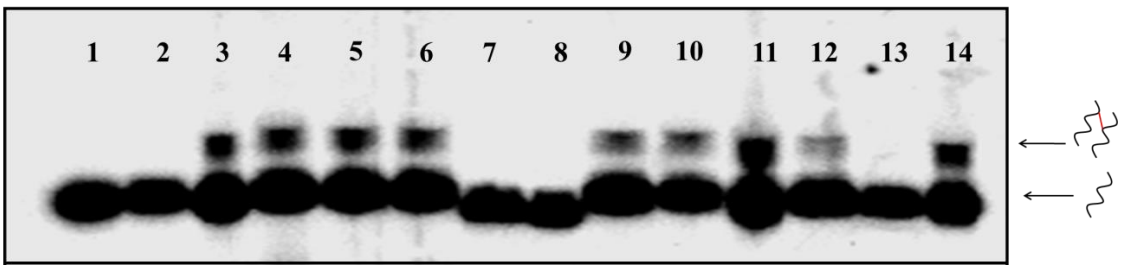
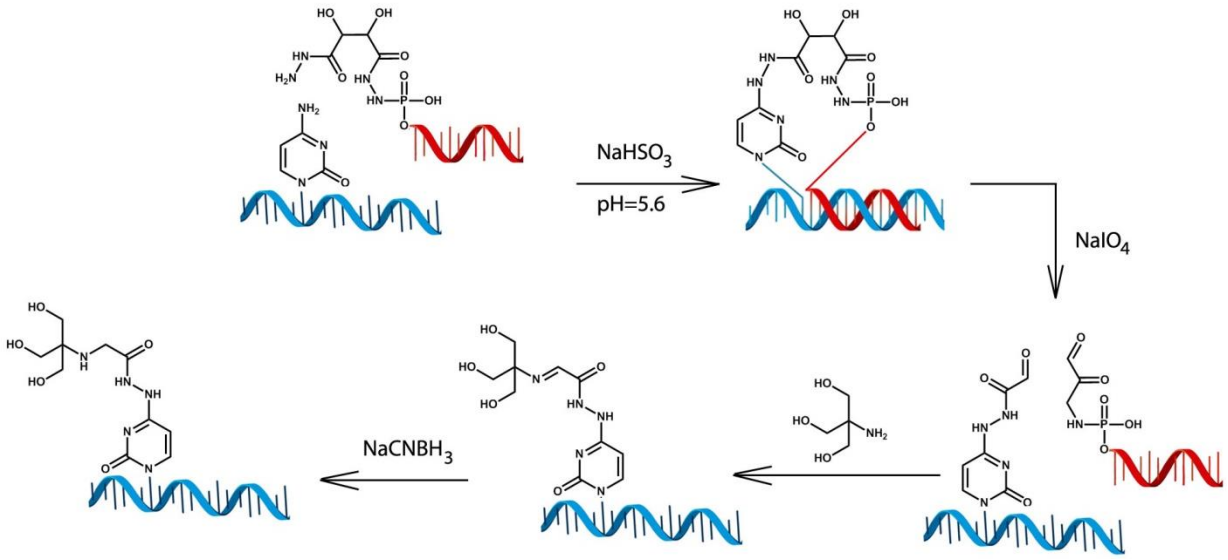


Figure 3.13. Cross-linking of 10-mer hydrazido-containing reactive sequences with complementary 13-mer target oligonucleotide with cytidine in the n+1 position (ODN_3.4) in ratio SRS:ODN = 2:1. Donor DNAs were modified with different linkers through 5’-cytosine (lanes 3-8, 14) or through phosphoramidate bond (lanes 9-13) and incubated with templates under transaminated conditions. After 3 days products with cleavable linkers were oxidized with sodium periodate (lane 7 and 13) and samples were resolved by 12% denaturing PAGE. As a positive control we used the unmodified 5’-cytosine recognition sequence (lane 1), the template which was treated under the same conditions (lane 2) and the amino-containing reactive sequence which was treated under the same conditions (lane 14).

3.7 Application of N4-transamination of cytosine for site-specific labeling.

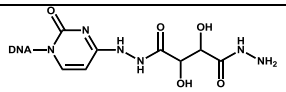
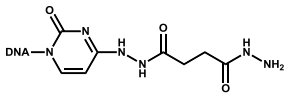
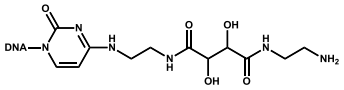
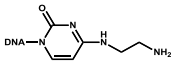
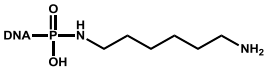
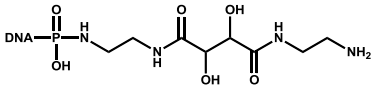
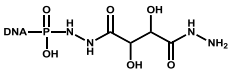
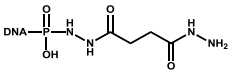
In the previous chapters we optimized the structure and reactivity of the recognition sequences containing cleavable linkers and identified that the hydrazine group demonstrated a higher activity for cross-link formation than the amino group. In the following step described in this chapter we oxidatively cleaved the respective linkers with $NaIO_4$ and tried to trap the newly formed aldehyde group. In control experiments we used constructs with noncleavable linkers and amino-reactive groups (**Scheme 3.8**).



Scheme 3.8. Transamination of cytosine, cleavage of linker with $NaIO_4$ and capture of the formed aldehyde group with tris(hydroxymethyl)aminomethane is presented.

A series of reactive sequences (**Table 3.4**) (SRS_3.26 – SRS_3.35) were prepared and hybridized with the complementary target oligonucleotide ODN_3.7 according to different transamination protocols. The reactive sequences contain two types of reactive groups: a primary amine (SRS_3.25 – SRS_3.28 or a hydrazine (SRS_3.26, SRS_3.27, SRS_3.32, and SRS_3.33) and various cleavable and non-cleavable linkers on the 5'-end. In the original protocol, i.e. protocol №1, the two DNA strands were annealed and the resulting duplexes were incubated for 3 days in presence of sodium hydrogensulfite (**Fig. 3.14, lanes 4-11**). Alternatively the target DNA strand was incubated with sodium hydrogensulfite only for 12 hours and then the reactive sequence was added, i.e. protocol №2 (**Fig. 3.14 lane 12-19**). After incubation for additional 12 hours the products were resolved with 12% denaturing PAGE.

Table 3.4. List of linkers and reactive groups that were investigated in cross-linking experiments[1] with the second generation modular system.

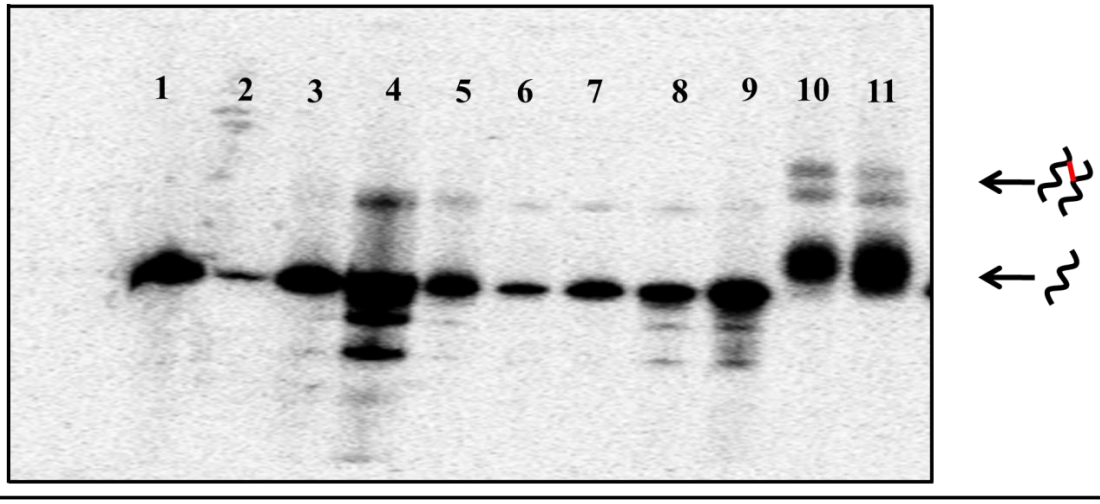
Entry	№ of compound	structure	Yield of synthesis SRS
1	SRS_3.26		89
2	SRS_3.27		92
3	SRS_3.28		60
4	SRS_3.29		98
5	SRS_3.30		92
6	SRS_3.31		75
7	SRS_3.32		67
8	SRS_3.33		79

1) Reactive sequence: 5’-(X)-CAA GTA GAA GAA GGG -3’. Target template contain cytosine core in position n+1 in overhang region; 5’- CCC TTC TTC TCA TTG **C**AA-3’ (ODN_3.7)

All eight of the incubated duplexes gave bands that have a reduced electrophoretic mobility, consistent with the formation of duplex DNAs containing interstrand cross-links. Higher yields were obtained when hydrazine derivatives were used as reactive groups. The two transamination methods, i.e. protocol №1 and №2, demonstrated similar efficiency in formation of cross-linked products.

A

№ line	1	2	3	4	5	6	7	8	9	10	11
№ compound	SRS_ 3.25	SRS_ 3.25	SRS_ 3.25	SRS_ 3.26	SRS_ 3.27	SRS_ 3.28	SRS_ 3.29	SRS_ 3.30	SRS_ 3.31	SRS_ 3.32	SRS_ 3.33
Amino-group						+	+	+	+		
Hydrazine-group				+	+					+	+
5'-Phosphate site								+	+	+	+
5'-Cytidine site				+	+	+	+				
Protocol №1		+		+	+	+	+	+	+	+	+
Protocol №2			+								
Yield in ICL formation, %	0.4	0.8	0.2	7	9	7	8	4	2.8	21	18



B

№ line	1	2	3
№ compound	SRS_ 3.25	SRS_ 3.25	SRS_ 3.25
Amino-group			
Hydrazine-group			
5'-Phosphate group			
5'-Cytosine group			
Protocol №1		+	
Protocol №2			+
Yield in ICL formation	0.4	0.8	0.2

12	13	14	15	16	17	18	19
SRS_ 3.26	SRS_ 3.27	SRS_ 3.28	SRS_ 3.29	SRS_ 3.30	SRS_ 3.31	SRS_ 3.32	SRS_ 3.33
		+	+	+	+		
+	+					+	+
				+	+	+	+
+	+	+	+				
+	+	+	+	+	+	+	+
6	7	2.4	2.2	3.6	1.2	17	19

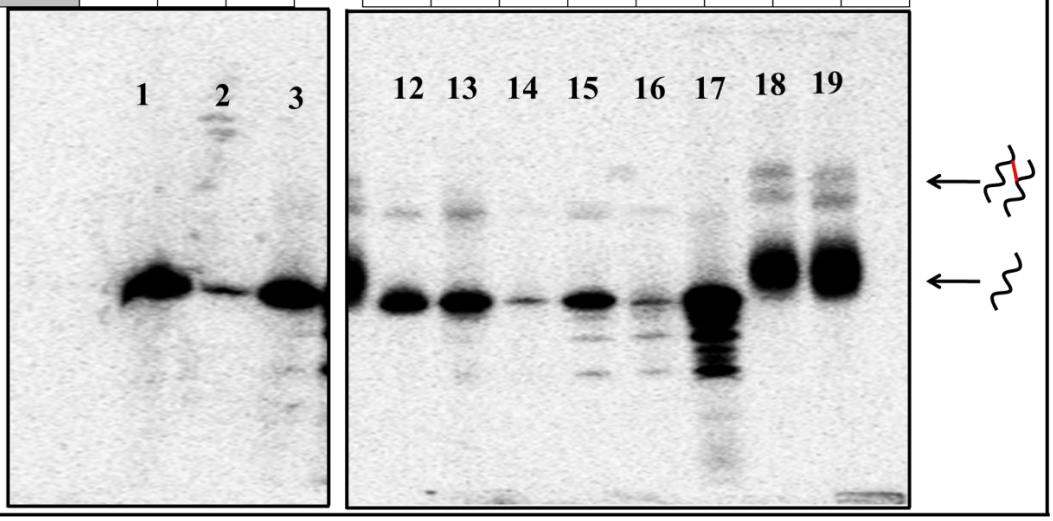


Figure 3.14. Cross-linking of 15-mer reactive sequences with a complementary 18-mer target oligonucleotide with cytidine in the position n+1 (ODN_3.7). Donor DNAs were modified with

amino- or hydrazido-containing linkers through 5'-cytosine (lanes 4-7, 12-15) or through phosphoramidate bond (lanes 8-11, 16-19). Cross-link reactions were carried out according to protocol №1 (lanes 4-11) or protocol №2 (lanes 12-19) As control we used the initial template ODN_3.7 (lane 1), the template which was treated under protocol №1 (lane 2) or under protocol №2 condition (lane 3).

To analyze the reactions by HPLC, we chose a reactive sequence with hydrazine reactive group and diol-containing linker (SRS_3.32). The complementary sequence (ODN_3.7) contained cytidine in position n+1 relative to the reactive group in the overhang region. The difference in retention times between reactive sequence and template allows us to follow product formation by reverse-phase HPLC chromatograms. After preactivation of the template according to protocol №2 and incubation of reactive and target strands for 12 hours, the HPLC chromatograms showed formation of a new peak at 13.5 min (**Fig. 3.15, trace 5**). After oxidation of the linker with $NaIO_4$, the formed aldehyde group was trapped with tris(hydroxymethyl)aminomethane with formation of an imine bond (Schiff base). Such a Schiff base is labile under neutral pH and can reversible hydrolyze to amine and aldehyde group. Hence in the last step the imine bond was selectively reduced with sodium cyanoborohydrate to the more stable secondary amine[20]. Accordingly, the peak at 13.5 min assigned to the cross-linked duplex DNA disappears after cleavage with $NaIO_4$. In contrast, the peak of the cross-linked product was still present after oxidation with $NaIO_4$ when the reactive sequence with a non-cleavable linker was used (**Fig. 3.15, traces 3** and **4**),

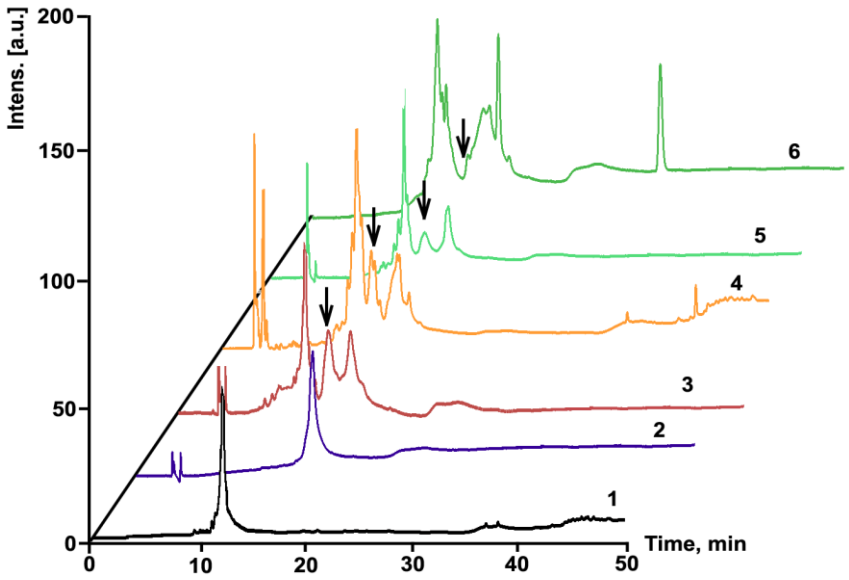


Figure 3.15. Reactions of hydrazide-containing reactive sequence SRS_3.32 (curves 5 and 6) and SRS_3.33 (traces 3 and 4) and complementary template ODN_3.7 that were transaminated according to protocol №2. RP-HPLC chromatograms (λ = 260 nm) demonstrated: **1:** black line – reactive sequence (SRS_3.33); **2:** violet line – template sequence (ODN_3.7); **3:** red line – cross-link formation between reactive sequence with non-cleavable linker (SRS_3.33) and template; **4:** orange line – oxidation of sample in trace 3 with $NaIO_4$; **5:** light green – cross-link formation between reactive sequence with cleavable linker (SRS_3.32) and template; **6:** dark green – oxidation of sample in trace 5 with $NaIO_4$. Black arrow indicates peak corresponded to cross-linked product.

Next we checked the possibility to trap the formed aldehyde group with additional probes. We used 5-(aminoacetamido)fluorescein as this probe contains a primary amino group for trapping of modified oligonucleotides and demonstrated stability during condition of reductive amination. Its absorption maximum at λ = 490 nm and strong fluorescence emission at λ = 514 nm provides tools to analyze its incorporation into the DNA strand using different types of spectroscopy. Again, first the cross-linked product between SRS_3.32 and ODN_3.7 was prepared using protocol №2. This sample was dialyzed overnight and precipitated with ethanol to remove hydrogensulfite ions which quenches $NaIO_4$. After purification, the oligonucleotide mixture was dissolved in 0.1 M sodium acetate buffer at pH = 5.6 and the diol linker was oxidized with $NaIO_4$ resulting in formation of two aldehyde groups. After incubation for 2 hours, the excess of oxidizing reagent was quenched with $NaHSO_3$ and the pH of the reaction was increased to 8.9 by addition of 1 M carbonate buffer. 5-(Aminoacetamido)fluorescein was added to the reaction to trap the aldehyde group by formation of a Schiff base. In the last step the imine bond formed between the fluorescein derivative and the modified template was reduced with

$NaCN_3BH_3$ to form a stable secondary amine bond. To analyze the final product we used RP-HPLC chromatograms at λ = 260 nm (**Fig. 3.16 A**). In a control reaction we labeled the reactive sequence SRS_3.32 after cleavage of the linker with $NaIO_4$ with 5-(aminoacetamido)fluorescein in the absence of target template ODN_3.7. RP-HPLC chromatogram demonstrated a series of peaks (from 8.1 min to 16.7 min) that correspond to unmodified DNAs as well as to product of modification of fluorescein with cleaved part of diol linker as well as formation of a sharp peak at 19.8 min assigned to labeled SRS_3.32. Comparing the control experiment with site-specific labeling of template we observed formation of an addition peak at 20.6 min that was assigned to labeled ODN_3.7. Analysis of reaction by fluorometry with excitation wavelength of λ = 492 nm and emission recording at λ = 514 nm show three peaks that correspond to unreacted fluorescein molecule (retention time 14.2 min), labeled SRS_3.32 at 19.8 min and labeled ODN_3.7 at 20.6 min. Position of unreacted fluorescein and labeled SRS_3.32 was identified by independent co-injections.

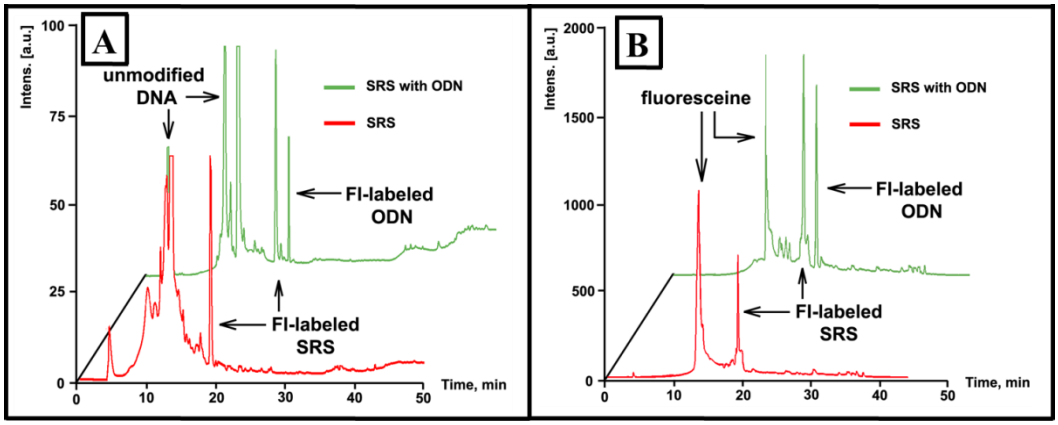


Figure 3.16. Reverse-phase HPLC chromatograms of labelling of aldehyde-modified DNA strand with 5-(aminoacetamido)fluorescein. **A:** Detection was done by UV spectrophotometry at λ = 260 nm. **B:** Detection was done by fluorometry with an excitation wavelength of λ = 492 nm and emission recording at λ = 260 nm. As a control, reactive sequence (SRS_3.32) was labelled under same conditions according to transaminated protocol №2.

3.8 “Locked” modular system

The stability of DNA duplexes is not only characterized by thermodynamic parameters but also by kinetic processes determining the speed of duplex formation and the time until unzipping. Even in DNA region, where only partial single strand formation occurs, a center of destabilization is created. This destabilization induced by missing neighboring base-pairs from one site leads to reduction of energy from π-π base stacking and hydrogen bond formation[21]. Results from the study of DNA helix unzipping processes indicate that some of the terminal nucleobases in duplex regions are not fixed within the duplex, but rather are in equilibrium with

a single stranded structure, which is called "DNA end-breathing"[22]. In our constructs, where the hypothesis of rigid DNA helix formation and localization of the reactive group close to the target site is a central dogma, this breathing can significantly decrease the effective concentration of the reactive group and hence decrease the yield.

To avoid this effect we used a construct, where the reactive group with cleavable linker is incorporated in the middle of the SRS. At each ends 10 complementary nucleotides create a locked construct. As previously mentioned, the hydrogensulfite-mediated labeling procedure is specific for cytosine in single-stranded nucleic acid regions. Hence a single-stranded loop region was additionally introduced into the locked construct by placing 2 nucleotide mismatches at either side of the cytosine carrying the reactive group to allow hydrogensulfite attack to the target cytosine in the template strand (**Fig. 3.17**).

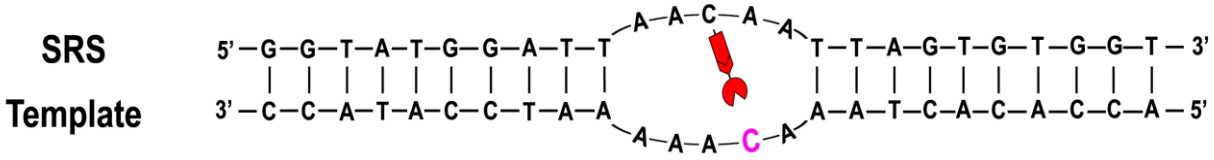
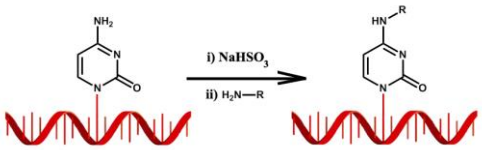
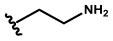
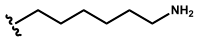
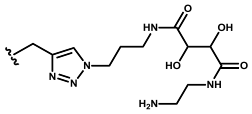
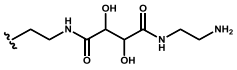
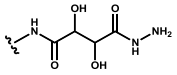
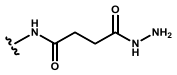


Figure 3.17. "Locked" modular system.

As it is possible that the linker can affect the hybridization efficiency, the duplex stability was evaluated with reactive sequences containing different linkers (SRS_3.35 - SRS_3.40) (**Table 3.5**). As expected a severe drop in stability can be observed, e.g. melting temperatures of 44 °C (SRS_3.36) and 36 °C (SRS_3.37) were measured compared to the experimental value of 48 °C for the duplex containing just a 5 central mismatches (**Fig. 3.18**). However, duplex stability is still sufficient for our experiments. To reduce the destabilization of the duplex observed for the triazol containing linker (SRS_3.37), the diol-building blocks SRS_3.38 – SRS_3.40 were additionally evaluated (the synthesis of building block is outlined in the previous chapter). The reactive sequences with different cleavable linkers were synthesized according to previously describe hydrogensulfite transamination of cytosine from recognition sequence and appropriate amine (or hydrazide). After synthesis molecules were dialyzed overnight and precipitated by ethanol. Reactive molecules SRS_3.37 – SRS_3.40 were additionally purified by RP-HPLC and all molecules were characterized by MALDI-TOF MS (**Table 3.5**)

Table 3.5. Characteristics of N4-amino-cytidine modified “locked” reactive molecules.



Entry	R	M_{calc}, Da	M_{found}, Da	Melting point[1]	Yield, %
SRS_3.35		7831.4	7832.2	47	95[2]
SRS_3.36		7887.4	7888.0	44	85[2]
SRS_3.37		8100.9	8096.2	36	45[3]
SRS_3.38		8005.56	8006.2	41	80[3]
SRS_3.39		7949.4	7950.0	45	60[3]
SRS_3.40		7917.4	7919.9	44	70[3]

1) Melting point was determined by measuring appropriate reactive molecule with complementary target sequence (ODN_3.8) at 100 mM sodium acetate buffer, pH = 7.2, 100 mM NaCl.
2) Synthetic yields were estimated by intensities of molecular ions peaks after MALDI-TOF MS experiment.
3) Synthetic yields were determined by integration peaks on RP-HPLC chromatograms.

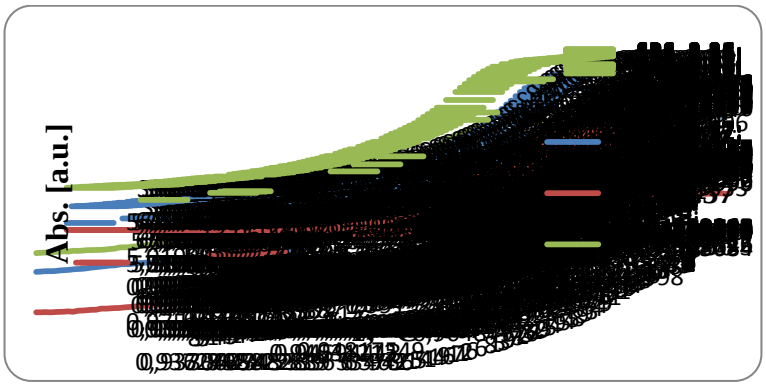


Figure 3.18. Comparison of melting curves for DNA template ODN_3.8 with **A**: SRS_3.35 **B**: SRS_3.36, **C:** SRS_3.37. Melting curves were recorded at λ = 260 nm with a heating rate of 1 °C/min. Duplex concentration was 2 mM in 100 mM NaCl and 10 mM sodium acetate buffer at pH 7.2.

Annealing of the reactive sequences SRS_3.35, SRS_3.36 and SRS_3.37 with the template sequence ODN_3.8 results in highly restricted flexibility of the nucleophilic reactive group in the loop and hence high reactivity towards cross-linking with the opposite cytosine base in the template sequence is observed, i.e. 21% compared to yields with the unlocked construct of 11%. (**Fig. 3.19, lanes 4-6**). Oxidation of cross-linked oligonucleotides with $NaIO_4$ results in cleavage of the linker in case of the diol-containing reactive sequence SRS_3.37 (**Fig. 3.19, lanes 10-12**) and formation of a highly reactive aldehyde group that could be potentially trapped with amino-derivatives in future experiments (**Fig. 3.19, lane 12**). Using hydrazine reactive groups (SRS_3.39 and SRS_3.40) cross-linked duplexes can be isolated in 60% yield (**Fig. 3.20**).

№ of lane	1	2	3		4	5	6		7	8	9	10	11	12
Template DNA	ODN _3.8	ODN _3.8	ODN _3.8		ODN _3.8	ODN _3.8	ODN _3.8		ODN _3.8	ODN _3.8	ODN _3.8	ODN _3.8	ODN _3.8	ODN _3.8
Donor DNA	SRS_ 3.35	SRS_ 3.36	SRS_ 3.37		SRS_ 3.35	SRS_ 3.36	SRS_ 3.37		SRS_ 3.35	SRS_ 3.36	SRS_ 3.37	SRS_ 3.35	SRS_ 3.36	SRS_ 3.37
$NaHSO_3$					+	+	+					+	+	+
$NaIO_4$									+	+	+	+	+	+
ICL yield, %	1.5	1.2	1.3		21	19	15		0.9	0.5	0.8	22	19	2.3

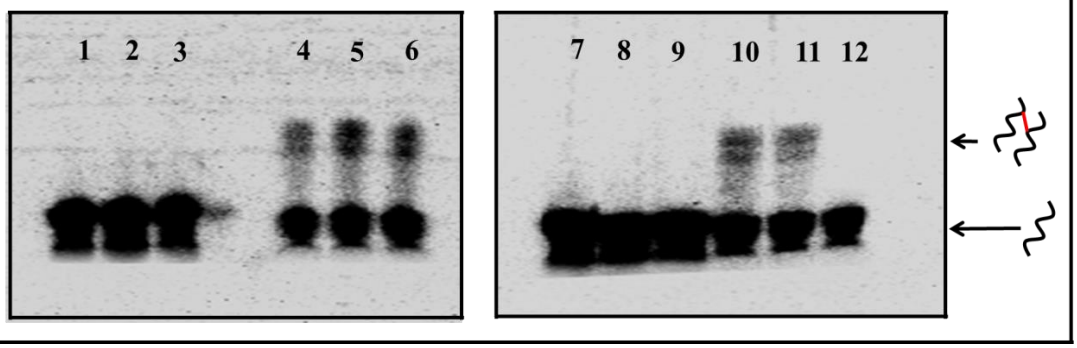


Figure 3.19. Denaturing PAGE analysis of DNA cross-link formation during hydrogensulfite catalysed transamination of cytidine using the locked constructs (lane 4-6, 10-12). Incubation of template (ODN_3.8) without activation with $NaHSO_3$ is shown in lanes 1-3, 7-9.

Based on our previous cross-linking experiment (**Fig. 3.14, Table 3.4**) the most efficient linker and reactive groups were tested in the "locked" constructs, i.e. SRS_3.36, SRS_3.38, SRS_3.39 and SRS_3.40. Sequences were incubated according to transamination protocol №1 (**Fig. 3.20, lanes 2-5**) or №2 conditions (**Fig. 3.20, lanes 6-9**) and yields of cross-link formation were determined by denaturing 12% PAGE. In correlation to our previous results, reactive sequences containing a hydrazine reactive group give significantly higher yields than amino-containing reactive sequences (**Fig. 3.21**). Additionally, both protocols demonstrated quiet similar yields in site-selective transamination of cytosine in template strand (ODN_3.8).

№ lane	1	2	3	4	5	6	7	8	9
№ of compound		SRS_3.36	SRS_3.38	SRS_3.39	SRS_3.40	SRS_3.36	SRS_3.38	SRS_3.39	SRS_3.40
Protocol 1	+	+	+	+	+				
Protocol 2						+	+	+	+
Yield in ICL formation, %	0.3	13	8	17	60	11	9	24	53

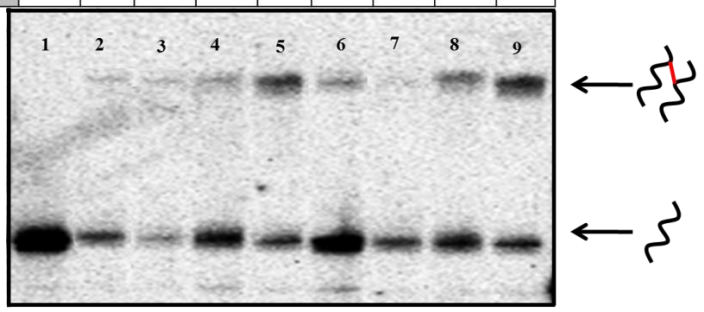


Figure 3.20. Denaturing PAGE analysis of cross-link formation during hydrogensulfite catalyzed transamination of cytidine using locked constructs with amino-containing reactive groups (lanes 2, 3, 6, 7) and hydrazine-containing reactive groups (lanes 4, 5, 8, 9) with complementary template ODN_3.8. Cross-linking experiments were done with protocol №1 (lane 2-5) and newly designed protocol №2 (lane 6-9). Incubation of ODN_3.8 without activation with $NaHSO_3$ is shown in lane 1.

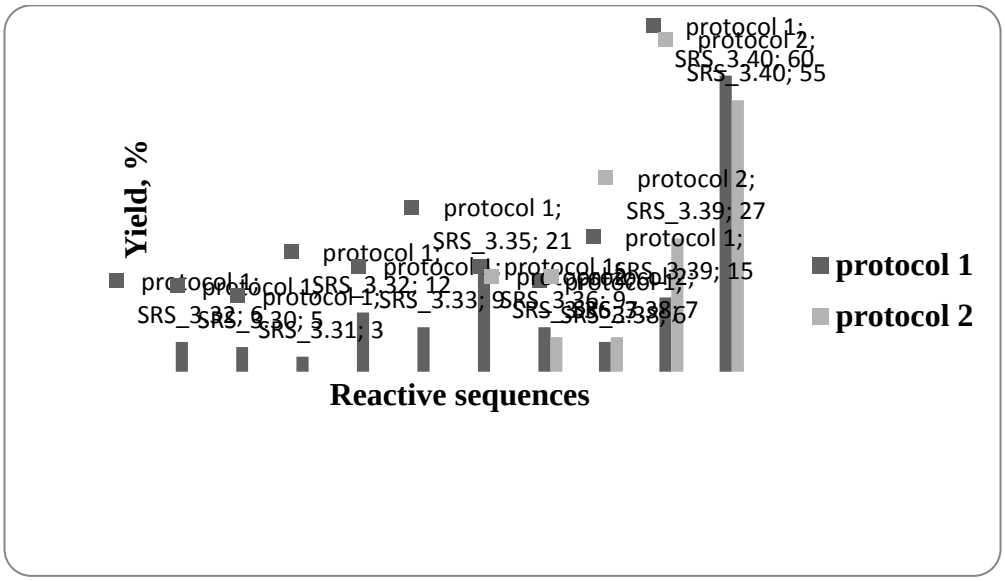


Figure 3.21. Yields of cross-link according to PAGE analysis (**Fig. 3.14** and **Fig. 3.20**) by determining the intensities of appropriate bands with the original software of the Typhoon Scanner PhosphorImager FLA 9500 with ImageQuant TL 1D (GE Healthcare).

3.9 Summary

In this chapter, we described the site-specific transamination of cytosine by single-step interaction between a nucleophilic reactive group of a reactive sequence and cytosine containing template DNA. Based on melting studies, our modular system forms stable duplexes with complementary single-stranded nucleic acids. The observed selectivity of the hydrogensulfite catalyzed reaction for cytosine represents a remarkable advantage. While most reactive sequences with primary amino-groups form cross-links to cytosine in the complementary sequence in low yield, hydrazine containing reactive groups exhibit significantly improved yields. The method described herein is one of the few strategies for site specific cross-link formation between two DNA strands.

The exact nature of the amino-containing reactive group used in this chapter has a considerable influence on the yield of cross-link formation. A delicate interplay between electronic and steric factors can be observed. The method is also versatile in this respect that a wide variety of other groups and markers e.g., chromophores, fluorophores, enzymes and electron-dense markers, can be attached when using cleavable linker that generate specific reactive group, e.g. an aldehyde, after cleavage. As example we have examined the labeling reaction of the aldehyde group, generated in the template strand after cleavage of the respective linkers, with a fluorescein derivative.

3.10 Reference

(1) Shapiro, R.; Weisgras, J. M. *Biochem. Biophys. Res. Commun.* **1970**, *40*, 839.
(2) Hayatsu, H.; Wataya, Y.; Kai, K. *J. Am. Chem. Soc.* **1970**, *92*, 724.
(3) Avignolo, C.; Valente, P.; Cai, S.; Roner, R.; Fulle, A.; Pizzorno, G.; Bignone, F. A. *Biochem. Bioph. Res. Co.* **1990**, *170*, 243.
(4) Sono, M.; Wataya, Y.; Hayatsu, H. *J. Am. Chem. Soc.* **1973**, *95*, 4745.
(5) Hayatsu, H. *J. Mol. Biol.* **1977**, *115*, 19.
(6) Avignolo, C.; Roner, R.; Cai, S.; Bignone, F. A. *J. Biochem. Bioph. Methods* **1991**, *23*, 193.
(7) Draper, D. E.; Gold, L. *Biochemistry* **1980**, *19*, 1774.
(8) Shapiro, R.; DeFate, V.; Welcher, M. *J. Am. Chem. Soc.* **1974**, *96*, 906.
(9) Shapiro, R.; Servis, R. E.; Welcher, M. *J. Am. Chem. Soc.* **1970**, *92*, 422.
(10) Molander, J.; Hurskainen, P.; Hovinen, J.; Lahti, M.; Lonnberg, H. *Bioconjugate Chem.* **1993**, *4*, 362.
(11) Hayatsu, H.; Wataya, Y.; Kai, K.; Iida, S. *Biochemistry* **1970**, *9*, 2858.
(12) Grimm, G. N.; Boutorine, A. S.; Helene, C. *Nucleos. Nucleot. Nucl.* **2000**, *19*, 1943.
(13) Viscidi, R. P.; Connelly, C. J.; Yolken, R. H. *J. Clin. Microbiol.* **1986**, *23*, 311.
(14) Lucas, R. L.; Benjamin, M.; Reineke, T. M. *Bioconjugate Chem.* **2007**, *19*, 24.
(15) Moquist, P. N.; Kodama, T.; Schaus, S. E. *Angew. Chem. Int. Ed.* **2010**, *49*, 7096.
(16) Gartner, Z. J.; Kanan, M. W.; Liu, D. R. *J. Am. Chem. Soc.* **2002**, *124*, 10304.
(17) Khullar, S.; Varaprasad, C. V.; Johnson, F. *J. Med. Chem.* **1999**, *42*, 947.
(18) Hayatsu, H. *Biochemistry* **1976**, *15*, 2677.
(19) Reisfeld, A.; Rothenberg, J. M.; Bayer, E. A.; Wilchek, M. *Biochem. Bioph. Res. Co.* **1987**, *142*, 519.
(20) Hermanson, G. T.; Bioconjugate Techniques, 2nd ed.; Elsevier Science: 2008, p 1323.
(21) Morales, J. C.; Kool, E. T. *Biochemistry* **2000**, *39*, 2626.
(22) Ambjörnsson, T.; Banik, S. K.; Krichevsky, O.; Metzler, R. *Biophys. J.* **2007**, *92*, 2674.

CHAPTER 4

Site-specific alkylation of guanine

4.1 Introduction

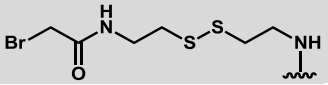
Functionalization of native ribonucleic and deoxyribonucleic acids has a paramount importance in all research fields related to life sciences. A variety of bioorthogonal groups and modifications in biomolecules are used for therapeutic delivery, targeted release, chemical reactions on molecules attached to a surface[1],[2], identification, sequencing and imaging of RNA[3], as well as for functional and structural studies of peptide-peptide, and oligonucleotide-peptide complexes[4],[5].

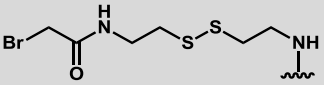
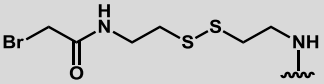
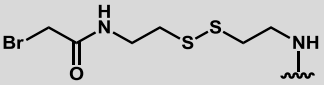
So far, site-specific reaction with nucleic acids focused mostly on the site-specific cleavage of nucleic acids[6],[7] or on DNA-templated chemistry to transfer reactive groups between two recognition sequences that are localized in close proximity to each other upon hybridization to a DNA template[8]. More recently, formation of cross-linked duplexes via a covalent bond was achieved[9]. The N^7 position of guanine is the most nucleophilic site among the heterocyclic bases of DNA[10]. Accordingly, when an electrophilic reactive group is positioned close in space to guanine, the predominant reaction involves covalent bond formation to this site. N^7-Alkyl-2'-deoxyguanosines were among the first covalent DNA adducts to be characterized[11]. However, application of guanine alkylation has been hampered, as most alkylated guanines undergo fast depurination resulting in an abasic site, which can lead to DNA strand breaks[12].

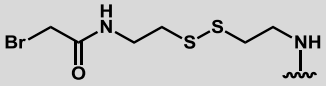
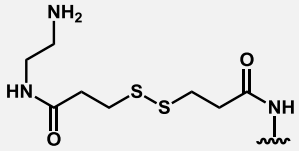
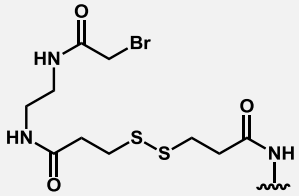
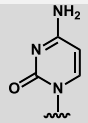
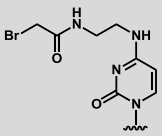
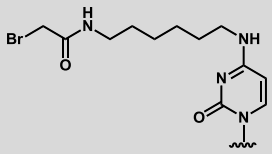
Nevertheless, this approach offers three major advantages. First, it allows to perform labeling of DNA under almost physiological conditions. Second, the reaction is performed in absence of additional reagents. Finally, as the N^7-position of guanine is located in a major groove alkylation can be performed in DNA duplex structures.

Sequences used:

* Complementary regions are underlined, (target G bases are indicated in red bold, SRS: reactive sequence; ODN: template; C* - cytidine in SRS that was modified with linker and reactive group in "locked" construct).

Name	Reactive group	Sequence
ODN_4.1	–	5'- CAC ACT TAT AA**G** AAA AAA AA -3'
SRS_4.1	PO_3H	5'- RG – <u>TTT ACA TCT C</u>-3'
SRS_4.2		5'- RG – <u>TTT ACA TCT C</u>-3'

ODN_4.2	–	5’- GAG ATG TAA A**G**T T -3’
ODN_4.3	–	5’- GAG ATG TAA AT**G** **G** -3’
ODN_4.4	–	5’- GAG ATG TAA A**G**T **G** -3’
ODN_4.5	–	5’- GAG ATG TAA AT **G**T -3’
ODN_4.6	–	5’- GAG ATG TAA ATT **G** -3’
ODN_4.7	–	5’- GCT CGC TCA A**G**T T -3’
SRS_4.3	PO_3H	5’- RG - TAT AAG TGT G -3’
SRS_4.4		5’- RG - TAT AAG TGT G -3’
SRS_4.5	PO_3H	5’- RG – TTT TTT TTA TAA GTG TG -3’
SRS_4.6		5’- RG – TTT TTT TTA TAA GTG TG -3’
ODN_4.8	–	5’- CAC ACT TAT A**G**A AAA AAA AA -3’
ODN_4.9	–	5’- CAC ACT TAT AA**G** AAA AAA AA -3’
ODN_4.10	–	5’- CAC ACT TAT AAA **G**AA AAA AA -3’
ODN_4.11	–	5’- CAC ACT TAT AAA A**G**A AAA AA -3’
ODN_4.12	–	5’- CAC ACT TAT AAA AA**G** AAA AA -3’
ODN_4.13	–	5’- CAC ACT TAT AAA AAA **G**AA AA -3’
ODN_4.14	–	5’- CAC ACT TAT AAA AAA A**G**A AA -3’
ODN_4.15	–	5’- CAC ACT TAT AAA AAA AA**G** AA -3’
ODN_4.16	–	5’- CAC ACT TAT AAA AAA AAA **G**A -3’
ODN_4.17	–	5’- CAC ACT TAT AAA AAA AAA A**G** -3’
SRS_4.7	PO_3H	5’- RG – TTT ACA TCT C-3’
SRS_4.8		5’- RG – TTT ACA TCT C-3’
ODN_4.18	–	5’- **G**A**G** AT**G** T**G**A ATA TCA TAT -3’
ODN_4.19	–	5’- **G**A**G** AT**G** TA**G** ATA TCA TAT -3’
ODN_4.20	–	5’- **G**A**G** AT**G** TAA **G**TA TCA TAT -3’
ODN_4.21	–	5’- **G**A**G** AT**G** TAA A**G**A TCA TAT -3’

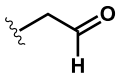
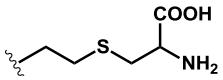
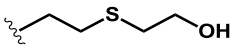
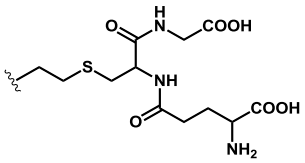
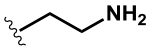
ODN_4.22	–	5’- **G**A**G** AT**G** TAA AT**G** TCA TAT -3’
ODN_4.23	–	5’- **G**A**G** AT**G** T**G**A ATA **G**CA TAT -3’
ODN_4.24	–	5’- **G**A**G** AT**G** T**G**A ATA T**G**A TAT -3’
ODN_4.25	–	5’- **G**A**G** AT**G** T**G**A ATA TC**G** TAT -3’
ODN_4.26	–	5’- **G**A**G** AT**G** T**G**A ATA TCA **G**AT -3’
ODN_4.27	–	5’- **G**A**G** AT**G** T**G**A ATA TCA T**G**T -3’
ODN_4.28	–	5’- **G**A**G** AT**G** T**G**A ATA TCA TA**G** -3’
RNA_4.1	–	5’ – GAG AUG UAA A**G**A U – 3’
SRS_4.9	PO_3H	5’-RG- TGG GGT GGG GTG GGT -3’
SRS_4.10		5’-RG- TGG GGT GGG GTG GGT -3’
ODN_4.29	–	5’- GGG GAG GGA GGG GAG GGG A**G**T AAA A -3’
ODN_4.30	–	5’- TTT TAC TCC CCT CCC CTC CCT CCC C-3’
PNA_4.1		5’- NH_2 – tat aag tgt g -3’
PNA_4.2		5’- X – tat aag tgt g -3’
PNA_4.3		5’- X – tat aag tgt g -3’
SRS_4.11		5’- GGT ATG GAT TAA **C***AA TTA GTG TGG T -3’
SRS_4.12		5’- GGT ATG GAT TAA **C***AA TTA GTG TGG T -3’
SRS_4.13		5’- GGT ATG GAT TAA **C***AA TTA GTG TGG T -3’
ODN_4.31	–	5’-ACC ACA CTA ATT **G**TT AAT CCA TAC C – 3’
ODN_4.32	–	5’-ACC ACA CTA **G**TT CTT AAT CCA TAC C – 3’
ODN_4.33	–	5’-ACC ACA CTA **G**TT **G**TT AAT CCA TAC C – 3’

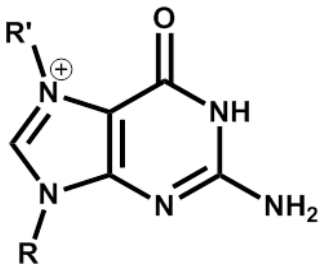
ODN_4.34	–	5’-ACC ACA CTA AAA **G**AA AAT CCA TAC C – 3’
ODN_4.35	–	5’-ACC ACA CTA **G**AA CAA AAT CCA TAC C – 3’
ODN_4.36	–	5’-ACC ACA CTA **G**AA **G**AA AAT CCA TAC C – 3’

4.2 Non-specific modification of DNA primer

Over the years, a series of guanine adducts have been synthesized and fundamental chemical properties of these lesions have been examined as well[13]. The results suggest that electron-withdrawing substitutions in the sugar ring[14] and in the N^7-side chain[15] can decrease the rate of depurination of guanine. After screening more than 200 published deoxyguanosine derivatives we selected side chain structures, which showed high stability under biological conditions (**Table 4.1**). For comparison, also side chain structures with low stability, i.e. few hours or less, are listed.

Table 4.1. Chemical stability of N^7-alkylguanine adducts.

R’	$T_{1/2}$, hours	Ref.
	3800	[16]
	62	[17]
	6.5	[18]
	52	[19]
	150	[20]
	0.15	[21]



Our general strategy includes recognition and annealing of the reactive sequence to the target DNA oligonucleotide, formation of cross-linked intermediates between reactive groups and target nucleobases and the subsequent cleavage of the covalent bond between the two DNA strands. We chose a cystamine derivative for the cleavable linker in our modular system that simply requires reduction with TCEP to cleave the disulphide bond and hence the cross-linked duplex and to form a new functional group, i.e. a thiol functionality. In the first step we tested

the stability of the N^7-guanine alkylated DNA strand under all possible conditions that we were planning to use in our reactions.

As a model reaction we performed electrophilic alkylation of a DNA primer (ODN_4.1.) which contains a single guanine base, with millimolar concentration of bromoacetic acid. The RP-HPLC chromatogram of the mixture after reaction demonstrated the presence of the same peak at 28.6 min after incubation for 12 hours. During MALDI-TOF MS measurements we observed the presence of two peaks with similar intensities (initial and modified DNA). Next, we incubated the mixture of modified and unmodified strand with the reducing agent TCEP to simulate cleavage of disulfide bonds. MALDI-TOF MS (**Fig. 4.1**) and RP-HPLC (**Fig. 4.2**) experiment demonstrated still one peak with retention time of 28.6 min with similar distribution in MS intensity between non-modified and modified DNA primers. In the last step we used piperidine cleavage to initiate DNA strand breaks at the alkylated guanine base and demonstrated formation of shorter DNA pieces eluting around 20.5 min in HPLC measurements. In the MS experiment the peak of modified DNA disappeared and a large number of signals for lower molecular weight fragments were observed.

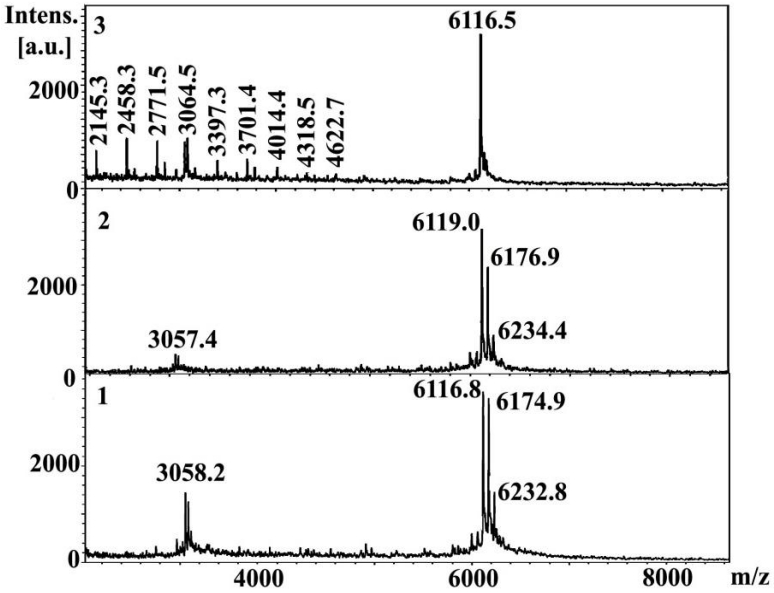


Figure 4.1. MALDI-TOF MS of N^7-(2-carboxy)-ethyl guanine oligonucleotide primer: **1**: ODN_4.1 after reaction with bromoacetic acid; **2**: after incubation of alkylated guanine DNA with TCEP; **3**: after incubation of alkylated guanine DNA with TCEP and piperidine cleavage.

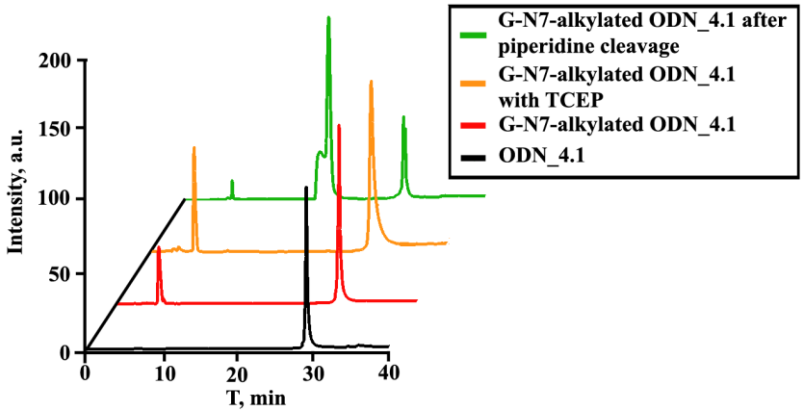
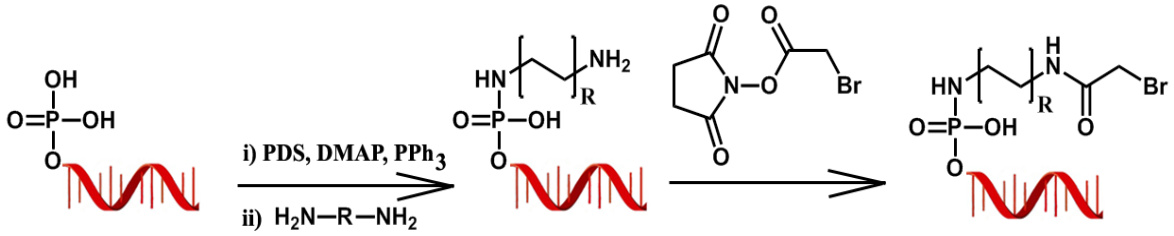


Figure 4.2. Reverse-phase HPLC chromatography of N^7-(2-carboxy)-ethyl guanine oligonucleotide primer: initial primer ODN_4.1 (black trace), ODN_4.1 after reaction with bromoacetic acid (red), incubation of alkylated guanine DNA with reducing agent (orange) and piperidine cleavage (green).

4.3 Site-specific modification of DNA primer: formation and selectivity

For the synthesis of the electrophilic reactive sequence the cleavable linker and the reactive group were prepared separately and coupled to the recognition sequence as shown in **Scheme 4.1** and described in the Experimental section. The general protocol was adopted with reference to Boutorine[22] or Knorre[23] protocols. The synthesis started from 5'-phosphate-containing DNA sequences with different length from 10 to 25 complementary nucleotides each forming stable duplex structures at room temperature with the target sequence (see list of used sequences). In the first step the phosphate-containing recognition sequence was reacted with diamino-containing linkers, i.e. ethylenediamine, hexylenediamine or cystamine, in presence of the coupling reagent PDS. The resulting modified DNA with a 5'-amino-group was obtained almost in quantitative yield. The acylation with N-hydroxysuccinimide ester of 2-bromoacetic acids in slightly alkaline buffer for 30 min gave the final reactive sequence in high yield (75% - 90%).



Scheme 4.1. Synthesis of reactive molecules by formation of phosphoramide bond and further acetylation with NHS-ester of 2-bromoacetic acid.

In the first trial we introduced cystamine to recognition sequence (SRS_4.1) as was describe above. Final coupling with 2-bromoacetic acid gave a reactive sequence in 89% yield. The reactive sequence SRS_4.2 with cystamine linker was reacted with ODN_4.2 as well as non-complementary sequence ODN_4.3 and cross-link formation was analyzed with 18% denaturing PAGE (**Fig. 4.3**). Formation of a faint band with lower electrophoretical mobility was observed for the reaction with ODN_4.2 in 0.5% yield but not for ODN_4.3.

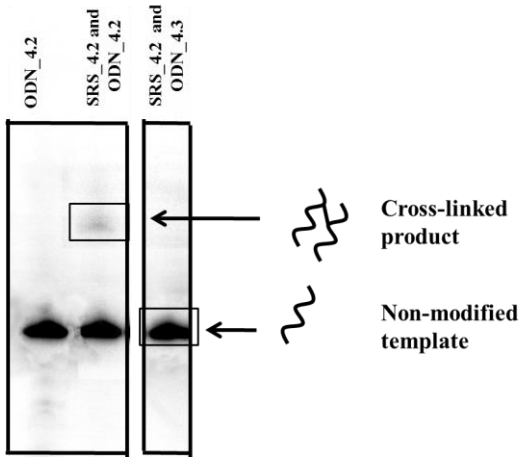


Figure 4.3. 18% Denaturing PAGE analysis of site-specific cross-link formation between 2-bromoacetate-containing reactive sequence SRS_4.2 and guanine containing complementary template ODN_4.2 (lane 2). As a control the non-complementary sequence ODN_4.3 gives no product (lane 3).

To optimize the labeling reaction time, temperature, as well as ratio between reactive and target sequences were varied and cross-link formation analyzed with PAGE. Increasing the ratio between reactive and template sequence from 1:1 to 2:1 and 10:1 increased the yield only 1.5 and 1.6 fold respectively. Hence, the ratio of 2:1 was chosen for future experiments. Next the incubation of reactive and target strand at different pH values from 5.6 to 9.3 for 16 hours demonstrated highest yield in the reaction with carbonate buffer at pH = 9.3 (6% to compare with 0.5% at pH = 7.4). Unfortunately, alkaline conditions also accelerate depurination of alkylated guanosine from DNA. Yields at pH = 7.4 and pH = 8.6 were significantly lower. i.e. 3% against 0.5%, but we decided to sacrifice a high yield for a more stable product at physiological conditions (pH = 7.4). In the next experiment we tested the influence of temperature on the efficiency of cross-link formation. We incubated reactive and target strand at +4°C that should stabilize the duplex and at +20°C as a control. Surprisingly, decreasing the temperature of incubation to +4°C not only stabilize the DNA duplex, but also lowered the yield of the reaction even further from 0.5% to almost 0.2%. Finally, we optimized the reaction time of guanine modification. After annealing we took aliquots after 5, 8, 12, 24, 48 and 72 hours.

Almost similar yields were demonstrated after 24 hours (0.5%) and 48 hours (0.7%) of incubation (we observed similar yield increasing with other type of constructs too, but differences were not more then 1.1 – 1.3 times with increasing incubation time in two times). We proposed that the low efficiency of cross-link formation was due to hydrolysis of the 2-bromoacetate group as well as to a parallel depurination of alkylated guanine. As a final conclusion, the incubation reaction was performed for 24 hours at +20°C with a ratio of reactive target sequence of 2:1 at pH = 7.4.

To evaluate the site-specificity of guanine alkylation SRS_4.4 containing the cleavable cystamine linker was reacted with a series of 10 templates, in which the position of single guanine was varied from positions (n+1) to (n+10) (ODN_4.8. – ODN4.17) relative to the reactive group of the reactive sequence. Cross-link formation was analysed with 18% PAGE (**Fig. 4.5**). As expected, after incubation the most reactive positions were the ones closest to the reactive group from the reactive sequence (ODN_4.8 and ODN_4.9). Yields decreased starting from the third position (ODN_4.10) and stabilize at 4% with G in (n+6) position and higher. Presence of cross-link bands with templates with G from (n+6) to (n+10) positions can be explained flexibility of short single strand DNA primer.

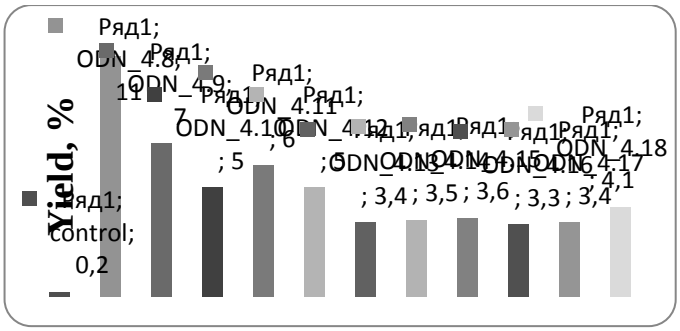


Figure 4.4. Efficiency of cross-link formation between SRS_4.4 and ODN_4.8 – ODN_4.17.

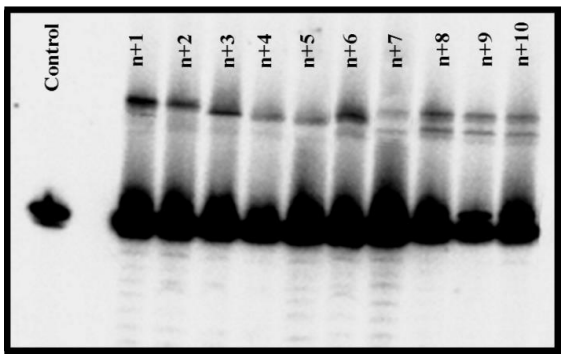


Figure 4.5. Phosphorimage autoradiogram of 18% denaturing PAGE analysis of alkylation of guanine with 10-mer reactive sequence (SRS_4.4), which contains a cleavable cystamine linker using complementary templates with G residues in different positions relative to the reactive group (ODN_4.8. – ODN_4.17). Incubation of template (ODN_4.8) in absence of reactive sequence is shown in lane 1.

These results nicely correlated with a simple mathematical evaluation. The theoretical model depicted in **Fig. 4.6** was based on the hypothesis, that reactive and target sequence form a stable duplex with rigid linear target DNA overhangs[24]. We compared the size of the reactive sequence (15 Å for linker and reactive group) with the distances between nucleosides in the target template (3.3 Å) and assumed that five nucleotides should be accessible for modification.

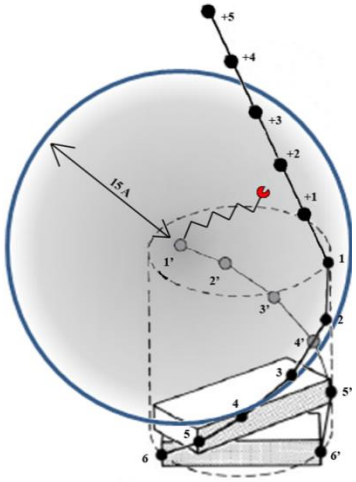


Figure 4.6. Simple mathematical approximation to evaluate dependence between linker length from reactive sequence and position of target nucleobase in target sequence. Figure was adapted from[24].

To examine the influence of the length of the complementary duplex and thus of the melting temperatures a longer 17-mer reactive sequence, i.e. SRS_4.6, which again contains a cystamine linker, was incubated with its complementary target strand ODN_4.15 with G in position (n+1). Increasing the number of complementary nucleobase in reactive and target sequence from 10 to 17 nucleobase led to increasing transition melting point form 34 °C (SRS_4.4 – ODN_4.8) to 53 °C (SRS_4.6 – ODN_4.15). In addition, as the N^7 position of guanine lies in the major groove and is not protected by formation of Watson-Crick hydrogen bonds we also tested two target sequences with G in position (n-2) (ODN_4.12) and (n-5) (ODN_4.9) (**Fig. 4.7**). As mentioned above, efficiency of cross-link formation was controlled by denaturing 12% PAGE analysis (in contrast to previous PAGE experiment, we used longer DNA primer that requires lower acrylamide percentages). The results show that the fraction of cross-linked product increases from 11% to 18% with increasing duplex stability from 34°C to 53°C for the reaction between SRS_4.6 and ODN_4.15. These results show that a 17mer length is preferable when oligoDNAs are used as hybridization probes. Hence, longer complementary sequences are crucial.

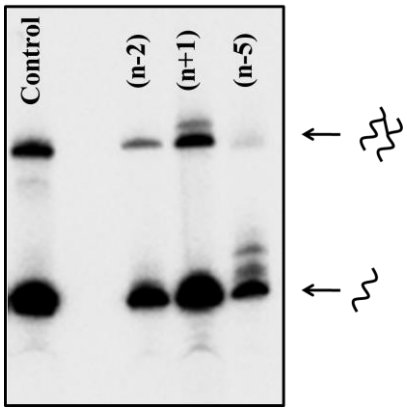


Figure 4.7. Phosphorimage autoradiogram of 12% denaturing PAGE analysis of alkylation of guanine with 17-mer reactive sequence (SRS_4.6), which contains a cleavable cystamine linker using complementary templates with G residues in (n-2) – ODN_4.12, (n+1) – ODN_4.15 and (n-5) – ODN_4.9. Incubation of template (ODN_4.8) with SRS_4.6 is shown in lane 1.

In the previous experiment we observed that modification of guanines in positions (n+1) to (n+3) is possible in moderate yields. Hence we tested templates that include two guanines in the "target region". The site-specificity of guanine alkylation was tested with piperidine cleavage of the strands at the position of depurinated modified guanine. We designed templates that include G in position (n+1) (ODN_4.2) (**Fig. 4. 8, lane 4**), (n+2) (ODN_4.5) (**Fig. 4. 8, lane 3**) and (n+3) (ODN_4.6) (**Fig. 4. 8, lane 6**). After modification with reactive sequence SRS_4.2 and piperidine cleavage all templates demonstrated formation of new bands that are different in their electrophoretic mobility. Then we checked selectivity with templates that include two guanine bases in position (n+1) and (n+3) (ODN_4.4) as well as G in position (n+2) and (n+3) (ODN_4.3). After site specific modification in case of ODN_4.4 (lane 2) we observed formation of two new bands with equal intensity. We assumed that almost similar efficiency in alkylation of guanine can be cause by high flexibility of last guanine in DNA sequence. At the same time, ODN_4.3 that contains two neighbouring G demonstrated preferential modification of G in (n+2) position compare to (n+3) position (lane 5). In control experiment we non-specifically alkylated ODN_4.3 with 2-bromoacetate and observed formation of two additional bands with desired electrophiretical mobility and similar intensities (**Fig. 4. 8, lane 1**).

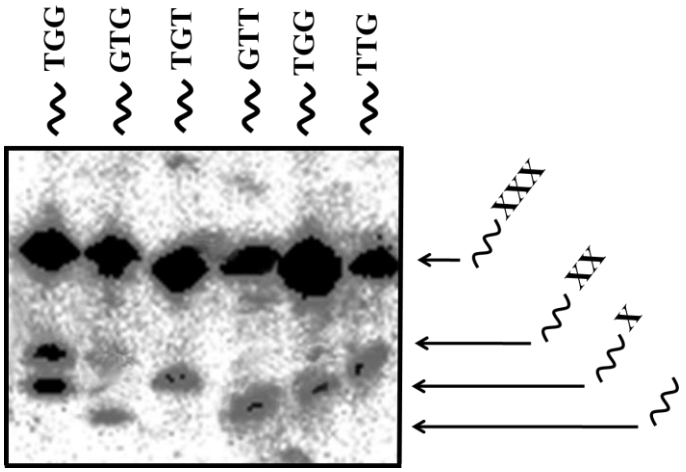


Figure 4.8. Determination of position of guanine alkylation by piperidine cleavage and analysis with 18% denaturing PAGE. In lane one template was non-specifically alkylated with 2-bromoacetate and cleaved by piperidine conditions.

4.4 Site-specific modification of DNA primers containing multiple guanines

We then proceeded to examine the alkylation of DNA templates with multiple guanine bases in the sequence. Sequences ODN_4.18 – ODN_4.22 contain three guanine bases each in the duplex region in positions (n-4), (n-7) and (n-9) and another guanine in a variable position from (n-2) (ODN_4.18) to (n+2) (ODN_4.18). Sequences ODN_4.23 – ODN_4.28 contain four guanine bases in the duplex region, i.e. (n-2), (n-4), (n-7), (n-9) and another guanine in a variable position from (n+3) (ODN_4.23) to (n+8) (ODN_4.28). Reaction of SRS_4.8 containing the cystamine linker with ODN_4.18 gives two cross-linked products, most likely with G in position (n-2) and (n-4). The cross-link with G in (n-2) position results in a T-shape structure under denaturating conditions, which is more bulky and hence has a lower electrophoretical mobility than the product of cross-link formation to G in position (n-4), which is more L-shape (**Fig. 4.9, lane 2**). When the guanine in the variable position gets closer to the reactive group (ODN_4.19 – ODN_4.21) the amount of L-shape product decreases and more T-shape products are formed (lanes 3 – 5). When the distance of the reactive group to both neighboring guanines becomes similar (ODN_4.22, lane 6) both products can be again observed. In ODN_4.23 an additional mismatches G in position (n-2) destabilizes the duplex and 3 products are observed probably involving the guanines in position (n+3), (n-2) and (n-4) (lane 7 and 8). In templates ODN_4.25 – ODN_4.28 guanine was placed almost in the end of template. In this construct, distance to G in the single strand region exceeds distance to G in double strand region. In the same time, reactive center is similar to ODN_4.18, in which G were placed in (n-2) and (n-4) positions. Similar reactive canter produce two similar upper ICL bands that are equal in electrophoretical mobility (lane 2 and lanes 9 -12). Third band can be corresponding to G in the single strand region.

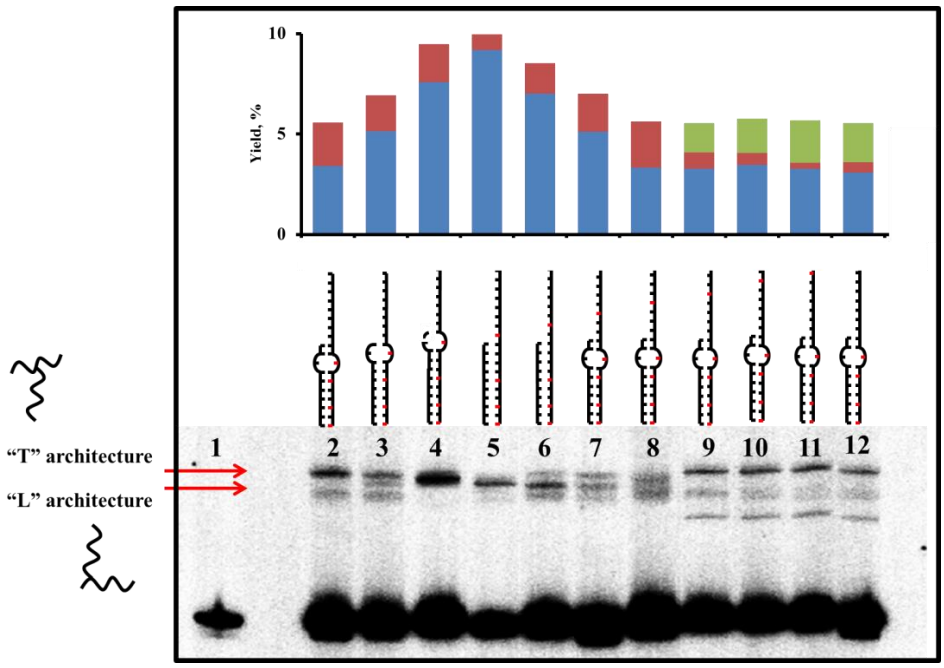


Figure 4.9. Phosphorimage autoradiogram of 18% denaturing PAGE analysis of alkylation of guanine between 10-mer reactive sequence SRS_4.8 and complementary templates with G residues in different positions (ODN_4.18 – ODN_4.28). Red dots mark the positions of guanine bases in template sequences. Incubation of the template in absence of SRS_4.8 is shown in lane 1. Yields of isolated cross-linked duplexes are indicated above the gel: blue bars indicate formation of T-shaped; red bars indicate formation of L-shape cross-linked products; green bars indicate formation of third unspecified cross-link product (probably – boundary L-shape architecture).

4.5 RNA template: point of interest

The genetic information of every organism is stored in form of deoxyribonucleic acid that is placed in different parts in the cell depending on the organism (prokaryotes, eukaryotes, bacteria or archae). Each DNA strand contains millions of specific regions that code for specific cellular functions and are called genes. For gene expression the enzymes RNA polymerase recognizes specific position on the DNA template and synthetizes new ribonucleic acid molecules (RNA) that are complementary to the respective region in the DNA. In the following steps the new RNA template is translated into a peptide sequence or protein by specific enzymes.

In the last decade, RNA has attracted additional interest and has gained evolutionary importance. It has been proposed that RNA both carries genetic information and serves as an enzyme[25]. Catalytic properties of RNA have been demonstrated on numerous examples such as Michael addition[5], peptide bond formation[26], nucleotide synthesis[27], etc. Based on this, it was hypothized that RNA might have been the starting compound for the evolution of life and the

term “RNA world” was introduced[28]. This hypothesis initiated a new wave of interest towards identification RNA structure and function in the living sell.

RNA consists of the nucleotides adenosine, guanosine, cytidine and uridine as monomeric building blocks. The structures of RNA mononucleotides is similar to DNA nucleotides but differ in the presence of an OH-group at position C2’ in sugar moiety. It was reviewed[14] that the electron withdrawing effect of this group influences the stability of the glycosidic bond and slows down the depurination reaction due to destabilization of the oxocarbenium ion in the transition step of the depurination reaction.

This ability of 2’-electron-withdrawing groups to slow down depurination suggests that the modular system might be also used for the site-specific labeling of RNA target templates. The results from DNA template modification suggest that guanine in the (n+1) position to the reactive group is modified with highest yield. We designed the RNA template RNA_4.1 accordingly. The three additionally present G bases in the duplex region should be sufficiently far away from the reactive group to be inert to modification. The cross-link reaction was performed with SRS_4.8 carrying the cystamine linker. As expected, the reaction proceeded without side reactions and demonstrated formation of only one additional band lower electrophoretical mobility in 6% yield on the 12% denaturing PAGE (**Fig. 4.10**).

Figure 4.10. Phosphorimage autoradiogram of 18% denaturing PAGE analysis of alkylation of guanine between 10-mer reactive sequence SRS_4.8 and a complementary template with G in the (n+1) position (RNA_4.1). Incubation of RNA_4.1 in absence of the reactive sequence is shown in lane 1.

4.6 Site-specific modification of guanine bases within the DNA double helix.

The position of the guanine N7-atom in the major groove of the DNA duplex and hence its accessibility make it a preferential target for modification. In addition, the sensitivity of alkylated guanine to alkaline conditions results in formation of DNA strand breaks. These properties inspired us to test the N7-position as a target in site-specific modifications. The discovery of triple-strand forming nucleic acids provided a tool for the recognition of specific region in the DNA duplex. Dervan at al[29] described that depending on the nucleotide content in the third strand, it can bind in parallel or antiparallel direction to the DNA duplex. In all motifs, the third strand attaches to the major groove resulting in an ideal steric localization of the reactive group close to the target N7-atom. In addition, the amount of purine and pyrimidine nucleotides in the sequence of the third DNA strand influences the stability of the triplex system. For example, cytosine-rich strands bind only under slightly acidic condition pH = 5.6[30] in contrast to guanine-rich strands which affinity is pH independent. Due to this fact we designed a double helix region with two templates (ODN_4.29 and ODN_4.30) and an appropriate third recognition sequence (SRS_4.10). The template ODN_4.29 contains a number of G bases in the duplex region as well as a target G in (n+1) position to the reactive group.

To test the site-specific alkylation of guanine, reactive sequence SRS_4.10 was prepared by the phosphoramidite method described in chapter 4.3.1 with cystamine linker and a bromoacetate derivative as a reactive group. SRS_4.10 forms a triple helix with the 25-mer duplex formed by ODN_4.29 and ODN_4.30. The cross-linking efficiency was tested under different conditions, i.e. in absence or presence of 5 mM Mg^{2+} ions and at +4 °C and +20 °C. Lower temperatures should stabilize the triplex helix, while the reaction should be faster at higher temperature according to the Van't-Hoff equation. An additional problem might be a certain amount of duplex formation between the G-rich reactive sequence and the C-rich ODN_4.30.

First we investigated triplex strand formation and stability with UV melting curves at λ = 260 nm in a buffer containing 100 mM NaC1 and 10 mM NaOAc, pH 7.4 (**Fig. 4.11**). The duplex formed by ODN_4.29 and ODN_4.30 shows a defined transition point at 74 °C, while the unmodified sequence SRS_4.9 neither forms a duplex with ODN_4.29 nor with ODN_4.30. The melting curve obtained for the equimolar mixture of all three strands shows a biphasic profile. The first transition point centered around 38°C is assigned to the triple helix formation of SRS_4.9 with the 25-mer duplex. The second melting point at 75°C nicely agrees with the melting temperatures of the 25-mer duplex determined before. The experiment clearly indicates that SRS_4.9 does not bind to the single strands but only to the 25-mer duplex.

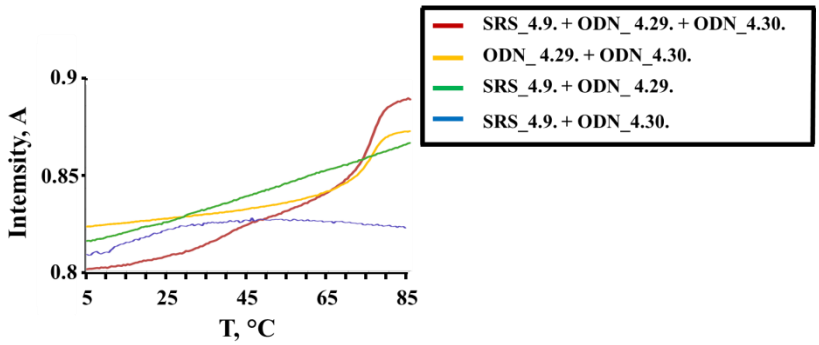


Figure 4.11. UV melting curves for different combinations of strands used for the triple helix-forming system. A mixture of SRS_4.9, ODN_4.29 and ODN_4.30 shows a biphasic profile with melting temperatures of 38°C and 75°C (red curve). T_m for the duplex formed by ODN_4.29 and ODN_4.30 is 74°C (yellow curve). No transition point is observed for the mixture SRS_4.9 and ODN_4.29 (green curve) or SRS_4.9 and ODN_4.30 (blue curve). Melting curves were recorded at λ = 260 nm with a heating rate of 1 °C/min. Duplex concentration was 2 mM in 100 mM NaCl and 10 mM acetate buffer at pH 7.4. SRS_4.9 concentration was 2 mM.

The efficiency of cross-link formation in the triple helix-forming system was examined with 12% denaturing PAGE (**Fig. 4.12**). Both target strand ODN_4.29 and auxiliary strand ODN_4.30 were labeled with [γ-^{32}P]-ATP. The two strands were mixed with ODN_4.10 in equimolar concentrations and heated to 85°C. After slowly cooling down for 3 hours, the reaction was incubated at neutral pH as described in the experimental section and **Fig. 4.12**. Both reactions at +4°C and +20°C give only very faint bands of the cross-linked product (lanes 3 and 4) with 0.3% and 0.7% percentage yields, respectively. In addition, the presence of three bands in total with reduced electrophoretical mobility indicate the formation of additional cross-linked products, probably involving the guanine bases in the (n-1) and (n-2) positions in ODN_4.29.

It is known that triplex motifs are strongly stabilized by the presence of divalent metal ions or multifunctional amines, which are thought to screen the charge interactions between the three negatively charged phosphodiester backbones[30]. In presence of 5 mM Mg^{2+} the yield of the reaction is increased to 0.6% at +4°C and even more pronounced to 3.9% at +20°C (lanes 5 and 6). In the control experiments (lanes 7 and 8), SRS_4.10 was annealed with the separate strands ODN_4.29 and ODN_4.30 in presence of 5 mM Mg^{2+}. In accordance with the UV melting studies, also here no cross-link formation is observed.

SRS_4.11				+	+	+	+		+	+
ODN_4.29	+			+	+	+	+		+	
ODN_4.30		+		+	+	+	+			+
T, °C				4	20	4	20		20	20
5mM Mg^{2+}						+	+		+	+
Yield				0.3	0.7	0.6	3.9		1.2	0.2

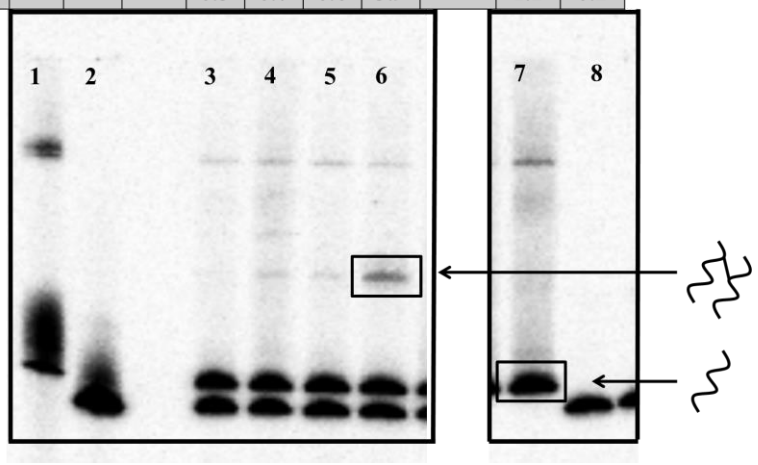


Figure 4.12. 12% Denaturing PAGE analysis of point-specific cross-link formation in triple helix-forming system. Line 1 – template ODN_4.29. This template contains a contamination with lower electrophoretical mobility that is unimportant for the experiment; line 2 – template ODN_4.30.; line 3 – reaction was incubated in absence of Mg^{2+} at T = +4 °C; line 4 – reaction was incubated in absence of Mg^{2+} at T = +20 °C; line 5 – reaction was incubated in presence of Mg^{2+} at T = +4 °C; line 6 – reaction was incubated in presence of Mg^{2+} at T = +20 °C; line 7 – reactive sequence SRS_4.10 was incubated with one template (ODN_4.29) at T = +20 °C in 5 mM Mg^{2+}; line 8 – SRS_4.10 was incubated with one template (ODN_4.30) at T = +20 °C in 5 mM Mg^{2+}. See explanation of reaction conditions in the text.

According to these results that showed the requirement for Mg^{2+} ions for successful cross-link formation, the Mg^{2+} concentration was optimized. Cross-linking formation between reactive sequence SRS_4.10 and the two templates ODN_4.29 and ODN_4.30 was repeated under similar conditions: ratio between reactive and template sequences was 2:1, reaction was incubated at +20°C for 24 hours with different concentrations of Mg^{2+} ions. The 12% denaturing PAGE demonstrated that increasing the Mg^{2+} concentration from 0 mM to 10 mM increased the yield of cross-link formation from 0.2% to 14%, respectively (**Fig. 4.13**). At this concentration, efficiency of cross-link formation reached its essential maximum, and further increase up to 100 mM had no effect. In addition, the band with lower electrophoretical mobility disappears after treatment with reducing agent that was correlated with cleavage of the disulfide bridge between reactive sequence and ODN_4.29 (**Fig. 4.13, lane 11**).

SRS_4.11			+	+	+	+	+	+	+	+	+
ODN_4.30	+		+	+		+	+	+	+	+	+
ODN_4.29		+	+		+	+	+	+	+	+	+
X mM Mg^{2+}	**5**	**5**	**5**	**5**	**5**	**1**	**10**	**20**	**100**	**0**	**5**
Yield, %			**2.7**			**5.2**	**14**	**12**	**13**	**0.2**	

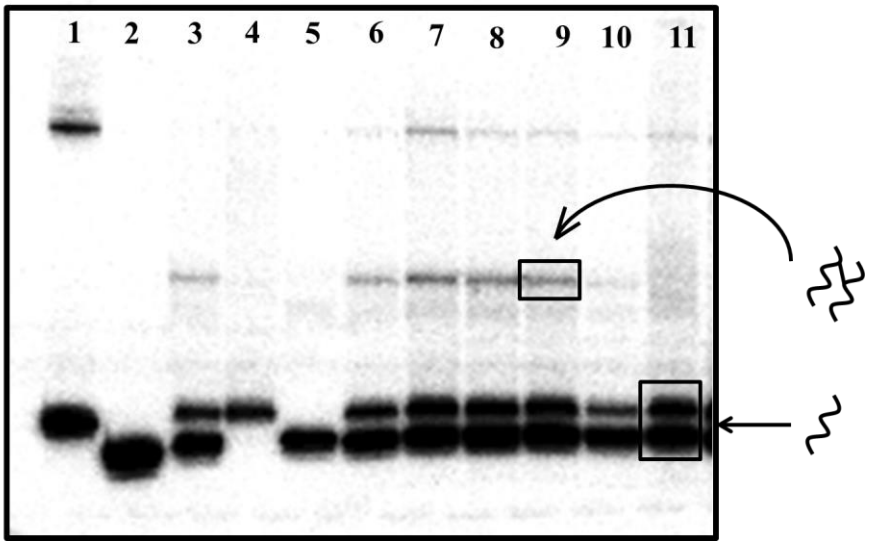
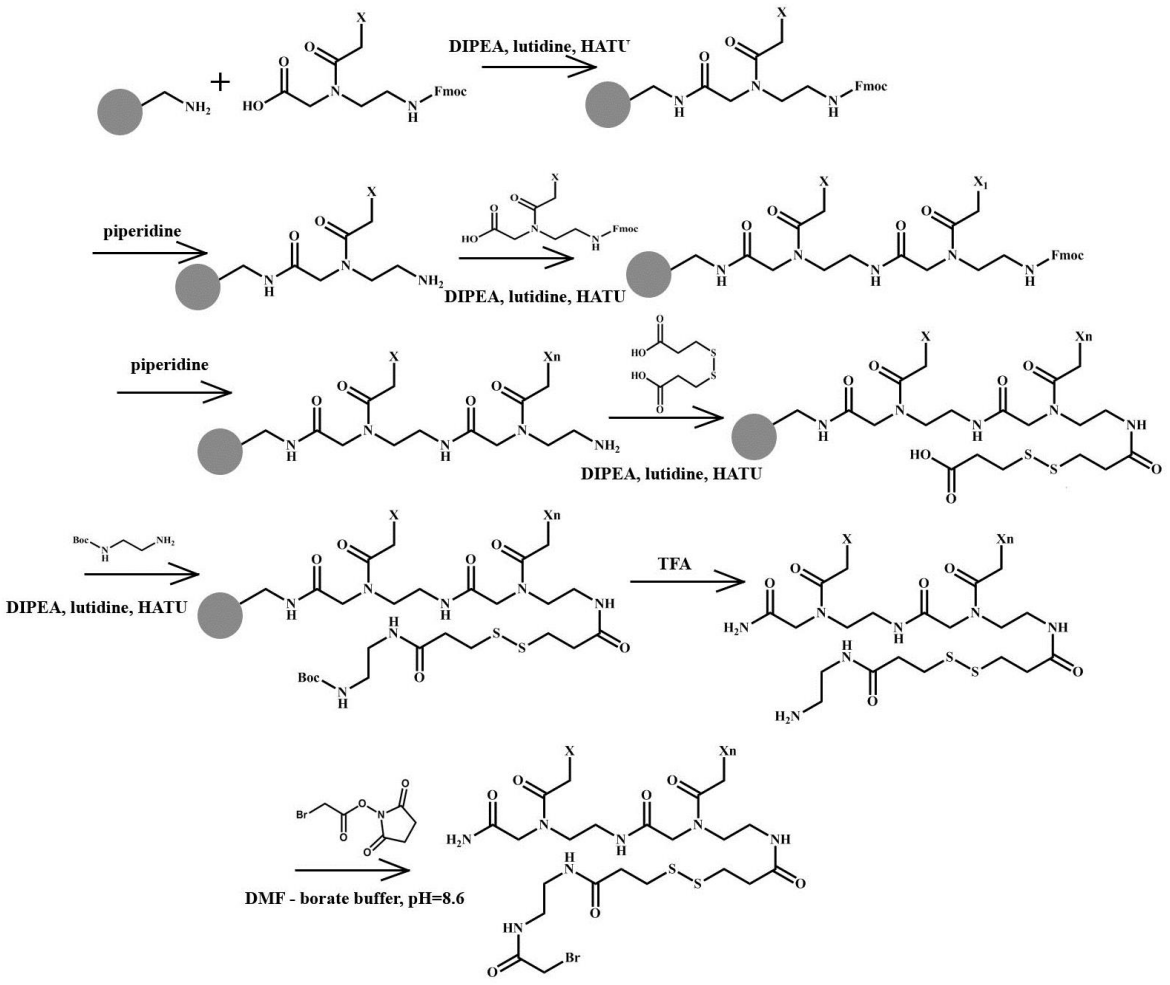


Figure 4.13. 12% Denaturing PAGE analysis of cross-linking efficiency in dependence of the Mg^{2+} concentration. Lane 1 – template (ODN_4.30.); lane 2 – template (ODN_4.29.); lane 3 – SRS_4.10. was incubated with ODN_4.29 and ODN_4.30 at 5 mM $MgCl_2$; lane 4 – SRS_4.10 was incubated with ODN_4.29 in 5 mM Mg^{2+}; lane 5 – SRS_4.10 was incubated with ODN_4.30 in 5 mM Mg^{2+}. SRS_4.10 was incubated with ODN_4.29 and ODN_4.30 at 1 mM $MgCl_2$ (lane 6); 10 mM Mg^{2+} (lane 7); 20 mM Mg^{2+} (lane 8); 100 mM Mg^{2+} (lane 9); 0 mM Mg^{2+} (line 10); 5 mM Mg^{2+} and additionally cleavage with TCEP (lane 11).

4.7 Use of PNA as recognition sequence

Binding affinity of DNA–DNA interactions and sequence specificity are two major characteristics of nucleic acid-based assays that determine their efficiency. Usually, these two properties are not correlated with each other. This lack of correlation means that, as affinity for the chosen target sequence increases, the likelihood of association with other similar target sequences also increases. Indeed, when the complementary parts of the reactive sequence and the target strand are long enough to form a stable duplex, the difference in energy between the completely matching duplex and a duplex with one mismatch approaches zero. Hence in DNA-template synthesis long recognition sequences could lead to recognition of the wrong sequence. To avoid this, the complementary sequence should be in a range, where one mismatch can still have an influence on the duplex stability but should be also long enough to recognize a specific sequence and to form a stable duplex. One of the possibilities to overcome this problem is the

use of peptide nucleic acids (PNA). PNA is an analog of ribonucleic acid consisting of monomers that also contain a heterocyclic ring (A, G, T or C) but in contrast to classical nucleotides these monomers are connected by the amino acid glycine. The size of the glycine backbone correlates with the size of the sugar ring that allows to place the nucleotides in a similar distance to each other as in natural nucleic acids[31]. Additionally, The presence of amide bonds instead of phosphate groups results in an uncharged PNA molecule and hence decreases the electrostatic repulsion between DNA and PNA strand in the duplex. This increases the melting temperature of the duplex and allows to shorten the complementary part of the reactive sequence.



Scheme 4.2. Solid phase synthesis of PNA reactive sequence PNA_4.3.

For this work, PNAs were synthesized by solid-phase peptide synthesis using Fmoc/Bhoc-protected PNA monomers and Fmoc-protected resin. The resin was deprotected with 20% piperidine in *N*,*N*-dimethylformamide and a 5-fold excess of PNA monomer in basic solution was loaded. For the initial monomer coupling to the resin the mixture was incubated for 2 h to

ensure complete coupling to the optimal number of resin sites. The resin was washed with DCM and DMF, and the whole process of deprotection and coupling was repeated until the desired PNA peptide was synthesized. For the introduction of a cleavable linker, we equipped the N-terminus of PNA, that corresponds to the 5'-end of DNA, with 3,3'-dithiopropionic acid as a cleavable block (**Scheme 4.2**). Next amide coupling between Boc-protected diaminoethane and the free carboxylic group followed by cleavage of PNA primer from resign resulted in formation of amino-modified PNA. In the last step 2-bromoacetic reactive group was included to yield the reactive sequence PNA_4.3 and cross-link experiments with ODN_4.8 were performed according to the previously described procedure. Quantification of the bands after 18% denaturing PAGE (**Fig. 4.14**) revealed efficiency of cross-link formation of 48%. This result showed that the highest duplex stability ($T_{melting\ point}$ = 68°C) resulted in the highest yield of modification obtained so far. It is important to note, that high stability of the PNA-DNA duplex might inhibit single strand formation under denaturing PAGE conditions. To avoid misinterpretation we performed a control experiment, where ODN_4.8 was annealed to PNA_4.2 that did not contain the 2-bromoacetate moiety. Accordingly, no cross-link formation was observed. Moreover we also did not observe the appearance of a band with lower electrophoretical mobility indicating the presence of non-covalent linking duplex.

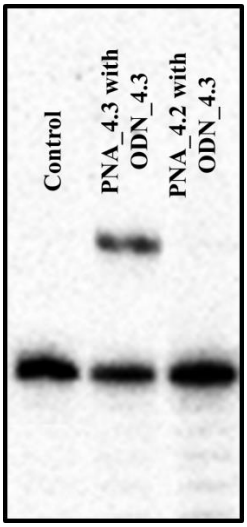


Figure 4.14. Phosphorimage autoradiogram of 18% denaturing PAGE analysis of alkylation of guanine between 10-mer reactive sequence (PNA_4.3) and complementary templates with G in (n+1) position (ODN_4.8). In control, we annealed the recognition sequence without reactive group (PNA_4.2) with ODN_4.8.

4.8 Correlation between yield of cross-link formation and duplex stability

We analyzed our previous results in order to confirm the dependence of cross-link formation efficiency and duplex stability. In our experiments we used reactive sequences ranging from 10 to 17 nucleobase in length resulting in melting temperatures between 32°C and 68°C. The highest melting temperature of 68°C was obtained with the PNA-based reactive sequence PNA_4.3. We observed a linear trend in yield increasing with increasing duplex stability. While the least stable duplex between SRS_4.2 and ODN_4.2 with a T_m of 32°C gave a yield of around 0.5%, the highest yield of 48% was obtained with the PNA-based system, i.e. T_m of duplex between PNA_4.3 and ODN_4.8 is 68°C.

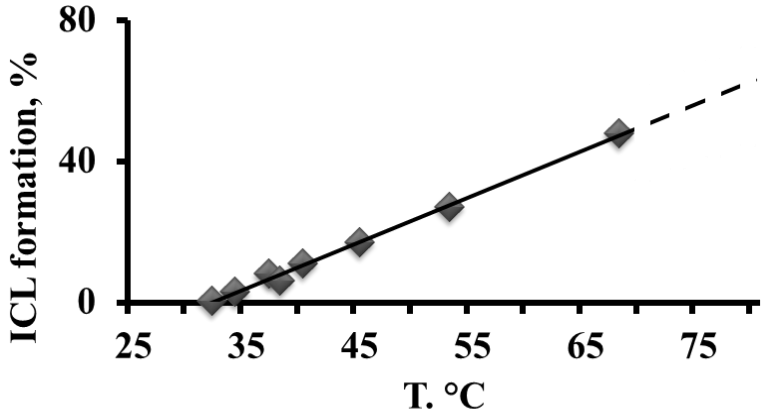


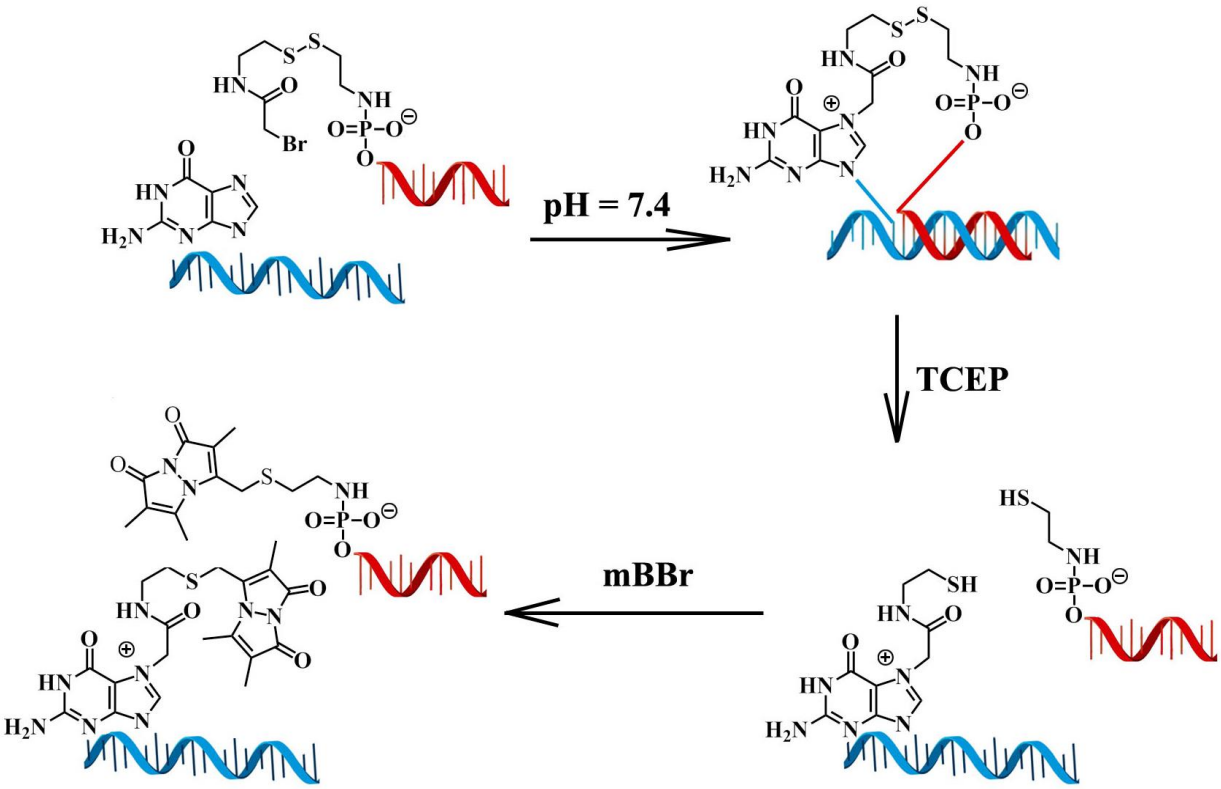
Figure 4.15. Graphical representation of the yield dependence of cross-link formation from duplex stability.

Table 4.2. Melting temperatures in °C and yields of cross-link formation for oligonucleotide duplexes.

Entry	Reactive sequence	Target sequence	T, °C	yield, %
1	SRS_4.2	ODN_4.2	32	0.5
2	SRS_4.8	ODN_4.18	34	3
3	SRS_4.8	ODN_4.21	37	8
4	SRS_4.2	RNA_4.1	38	6
5	SRS_4.4	ODN_4.8	40	11
6	SRS_4.6	ODN_4.15	53	27
7	PNA_4.3	ODN_4.8	68	48

4.9 Specific labeling of long oligonucleotides with fluorophores.

To further validate the modular system as a general tool for the site-specific labeling of long native oligonucleotides we attempted to couple a fluorophore to the thiol group obtained after reduction of the disulfide bridge containing cystamine linker. To test this approach we used bromobimane (mBBr) as a thiol reactive fluorophore (**Scheme 4.3**). There are several advantages of using mBBr as a probe for labeling our construct. Specifically, the small size of the molecule allows it to attach to essentially any position of an oligonucleotide structure avoiding restrictions associated with secondary and tertiary oligonucleotide structures.



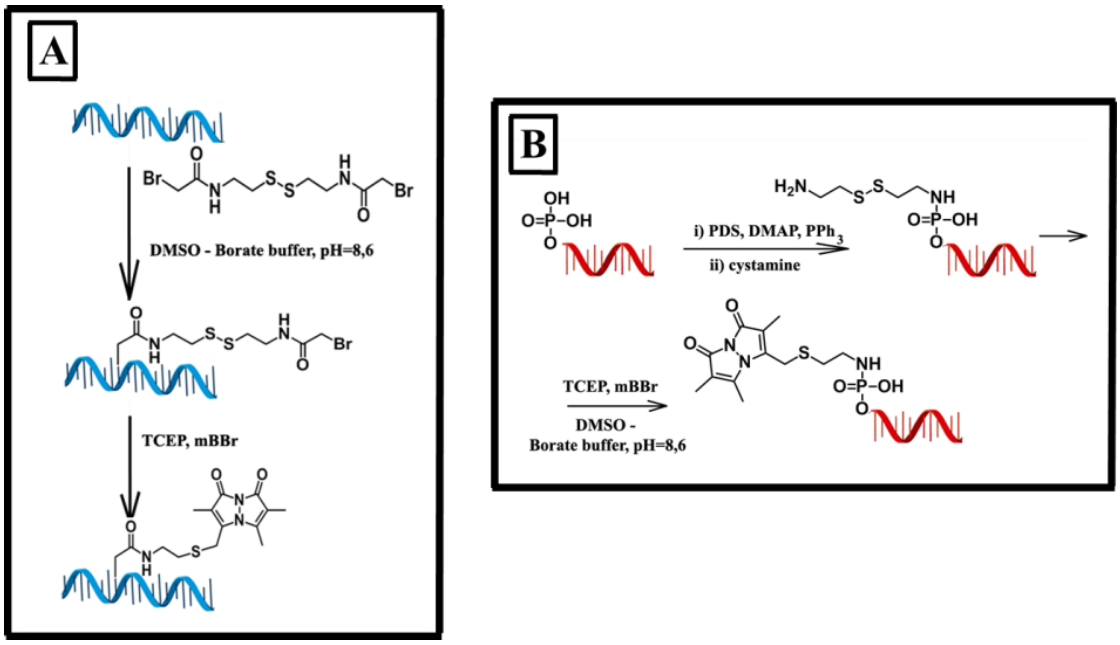
Scheme 4.3. Site-specific modification of target DNA strand (blue, ODN_4.2) with modular system (red, SRS_4.2). The newly introduced thiol group in the target oligonucleotide is labelled with the mBBr fluorophore after reduction with TCEP.

The synthesis started from the commercially obtained 5'-(PHOS)-DNA sequence (SRS_4.1). The cleavable block was introduced as cystamine moiety by phosphoramidite coupling in presence of PDS. After addition of hydroxysuccinimide ester of 2-bromoacetate the reactive sequence was precipitated with ethanol and directly annealed with template ODN_4.2. After incubation for 12 hours the formed cross-linked duplex was cleaved *in situ* with TCEP, and incubated with mBBr. The obtained oligonucleotide mixture was analyzed by HPLC and demonstrated formation of two groups of peaks at 28 and 32 min, respectively (**Fig. 4.16**). The

peak at 28 min shows only absorption at λ = 254 nm and has a retention time corresponding to the initial oligonucleotides. This assignment was verified by co-injection of SRS_4.2 and ODN_4.2 with the crude mixture. In addition, matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) of the peak revealed the presence of mass peaks corresponding to the masses of the unmodified strands ($M_{SRS\_4.2}$= 3031.5 Da; $M_{ODN\_4.2}$ = 4069.0 Da) (see **Fig. 4.17 A**).

The broad peak centered around 31 min and 33 min shows absorption both at λ = 260 nm, which originates from the oligonucleotide moiety, and at λ = 390 nm originating from the mBBr. Additionally, analysis of these peaks by MALDI-TOF MS revealed masses of 3279.6 Da and 4370.1 Da respectively (see **Fig. 4.17 B** and **C**). These results match very well with the calculated value of 3281.1 Da for SRS_4.2 and 4369.1 Da for ODN_4.2, both modified with on mBBr moiety. In the control experiment, SRS_4.2 was reduced and labeled in the absence of ODN_4.2 (**Scheme 4.4**). Co-injection of the control experiment solution with previously labeled mixture demonstrated peaks that overlapped at 31 min. Intriguingly, excitation of both HPLC peaks at λ = 390 nm demonstrated strong emission at λ = 475 nm which is in line with reference spectra of mBBr (**Fig. 4.18**).

To further verify formation of mBBr-labeled sequences the mixture after labeling was also analysed with 18% PAGE (**Fig. 4.17 D**). For comparison also the separately synthesized sequences (**Scheme 4.4. A** and **B**) were subjected to PAGE (mBBr-labeled SRS_4.2, lane 1; mBBr-labeled ODN_4.2, lane 3). A fluorescent scan of the gel shows two bands for the reaction mixture. Comparing migration of these bands with the two control lanes allows to assign the lower band to mBBr-labelled SRS_4.2 while the upper band corresponds to mBBr-labeled ODN_4.2).



Scheme 4.4. **A**: Modification of ODN_4.2 containing a single guanine in a non-specific reaction. **B**: Synthesis and 5'-labelling of SRS_4.2 with mBBr fluorophore.

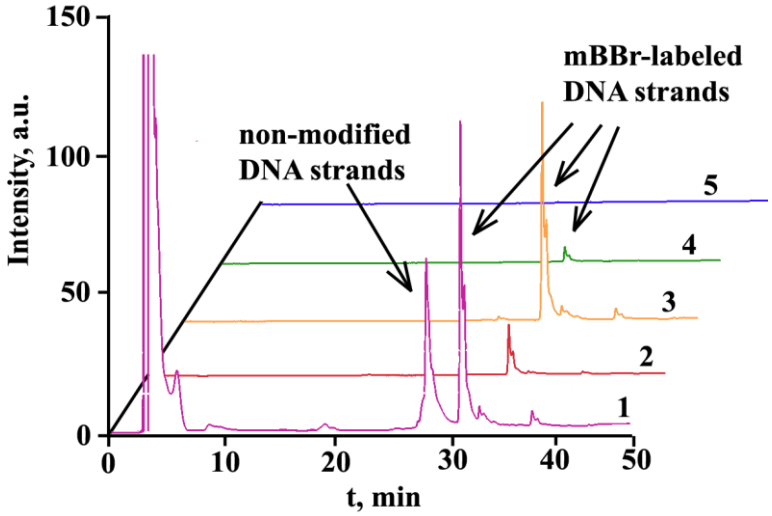


Figure 4.16. Reverse-phase HPLC chromatogram of mBBr labeled mixture obtained after reaction between SRS_4.2 and ODN_4.2 followed by reductive cleavage of the cystamine linker: **1**: λ = 272 nm, **2**: λ = 331 nm, **3**: λ = 390 nm, **4**: λ = 450 nm, **5**: λ = 550 nm.

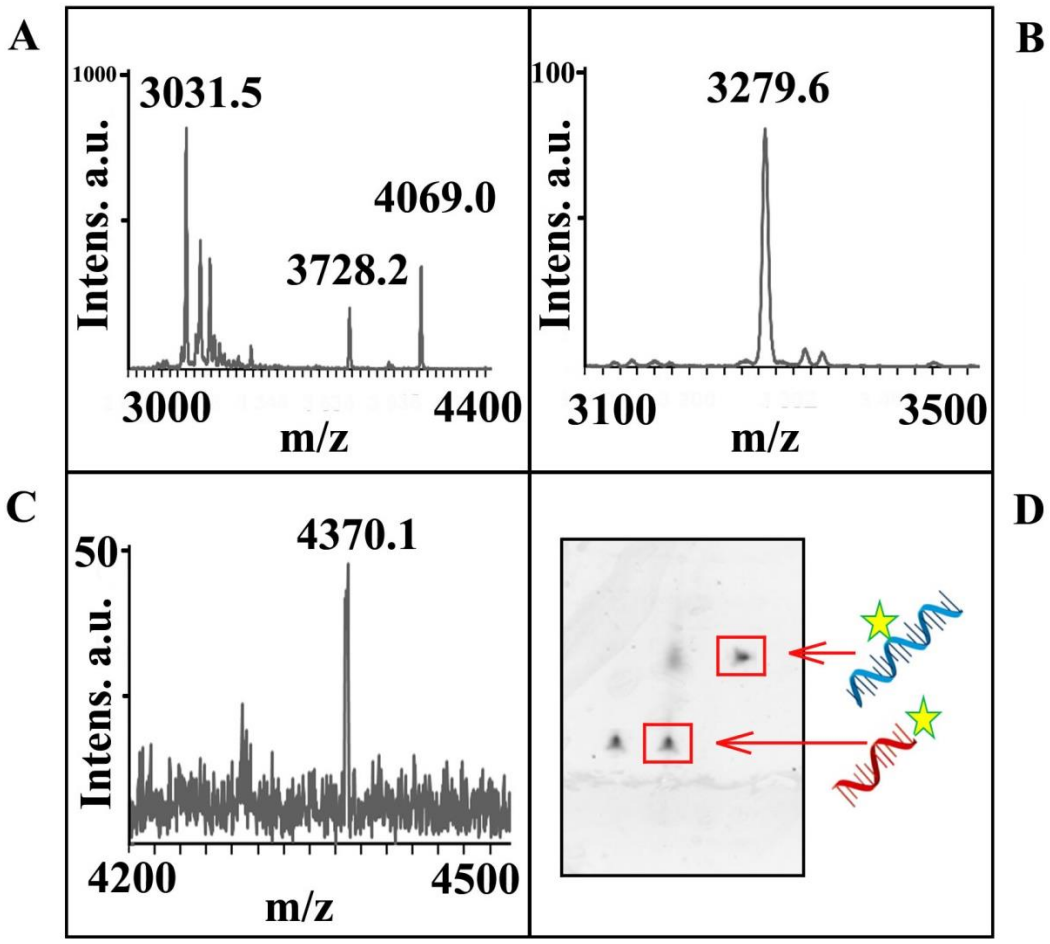


Figure 4.17. MALDI-TOF spectrum of peaks after RP-HPLC separation: **A**: MS of peak at 28 min demonstrated presence of nonmodified sequences: SRS_4.2 – $M_{exp.}$ = 3031.5 Da, $M_{calc.}$ = 3030.2 Da; ODN_4.2 after depurination of one G base – $M_{exp.}$ = 3728.2 Da, $M_{calc.}$ = 3729.8 Da; nonmodified ODN_4.2 – $M_{exp.}$ = 4069.0 Da, $M_{calc.}$ = 4070.1 Da. **B**: MS spectrum of peak at 31 min – mBBr-labeled SRS_4.2 ($M_{exp.}$ = 3279.6 Da, $M_{calc.}$ = 3278.6 Da); **C**: MS spectrum of peak at 33 min – mBBr-labeled ODN_4.2 ($M_{exp.}$ = 4370.1 Da, $M_{calc.}$ = 4374.0 Da); **D**: Fluorescent scan of a 18% denaturing gel, showing analysis of DNA labeling: lane 1 – 5'-labeling SRS_4.2 without complementary sequence; lane 2 – mBBr labeled mixture obtained after reaction between SRS_4.2 and ODN_4.2 followed by reductive cleavage of the cystamine linker; lane 3 – non-specific mBBr-labeled ODN_4.2. The mBBr-labeled SRS_4.2 and ODN_4.2 sequences are indicated on the right-hand side of the image.

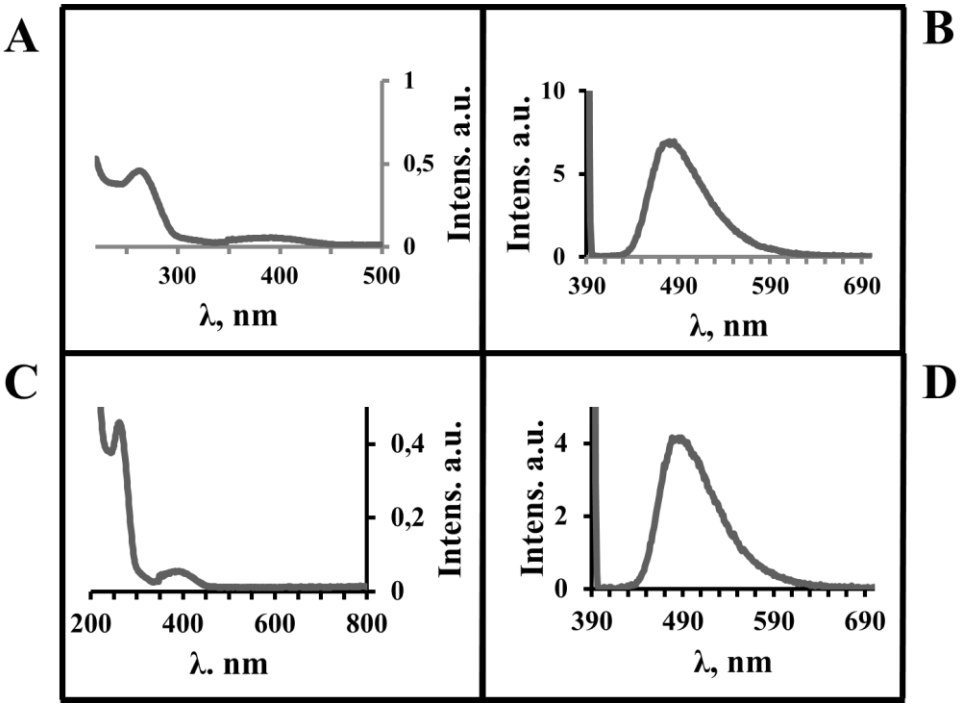
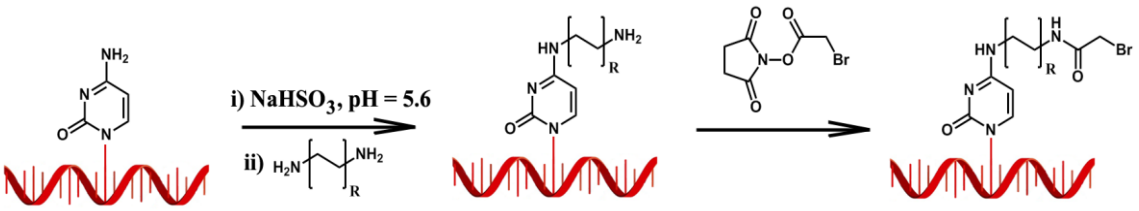


Figure 4.18. **A:** UV spectrum of peak at 33 min after RP-HPLC purification. **B**; Fluorescence spectrum of peak at 33 min after RP-HPLC purification. **C:** UV spectrum of peak at 31 min after RP-HPLC purification. **D**; Fluorescence spectrum of peak at 31 min after RP-HPLC purification;

4.10 "Locked" modular system

The transamination experiment with cytosine showed that the incorporation of a reactive group in the middle of the donor molecule and the stabilization of the reactive site by flanking duplex regions results in increased cross-link formation yields. To test this approach in guanine alkylation, we used SRS_4.11 with a single cytosine in the center of the sequence, which was transaminated with diamino-derivatives. Following aminoacylation with N-hydroxysuccinimide-2-bromoacetic acid, the SRS_4.12 or SRS_4.13 was mixed with appropriate templates, i.e. ODN_4.31 – ODN_4.36, under original conditions. As linkers we used diaminoethane and diaminohexane moieties. We assumed that two linkers with different length will demonstrate different flexibility of the reactive group in the duplex strand region resulting in different sterical arrangements of the reactive group relative to the target N7-atom of guanine in the major groove. From both ends, the reactive group is locked by ten complementary base pairs to reduce destabilization by flipping ends. The five nucleobases in the middle are designed to form either a duplex (ODN_4.31 – ODN_4.33) or a loop (ODN_4.34 – ODN_4.36).



Scheme 4.5. Synthesis of reactive sequences SRS_4.11 – SRS_4.13 as "locked" modular system.

For a more detailed evaluation of the importance of the linker size, target guanines were placed opposite to the reactive group (ODN_4.31 and ODN_4.34) or in the (n+3) position (ODN_4.32 and ODN_4.35). Additional template strands were designed to compare the reactivity of two guanines, which were placed opposite the reactive group and in the (n+3) position (ODN_4.33 and ODN_4.36).

In the currently studied crosslinking context, where the reactive group is located in the center of the oligonucleotide strands, the observed selectivity for cross-linking to different guanine positions can be possibly explained in terms of a major groove location of the target N^7 guanine atom and/or size of the linker in the reactive group. Analysis of the complete series of duplexes however, indicated formation of mainly one single cross-linked species in all cases as derived from PAGE experiments (**Fig. 4.19**). With completely complementary sequences only reactive sequence SRS_4.12 with diaminohexane linker demonstrated formation of an intensive band corresponding to cross-linked product. In contrast, in the experiment with the shorter diaminoethane linker no additional band was observed. In case of a fully complementary strand (ODN_4.31) the reactive group can only attack the target guanine by reaching around the helix. Only the diaminohexane linker is long enough to attack the N7 position in the major groove on the opposite site of the helix. The diaminoethane linker is too short and hence no cross-linked product is observed. The same effect is observed for ODN_4.33. Also here, the region, where the target guanine in position (n+0) is located, is completely complementary to the reactive sequence, while a single mismatch is present at the second guanine in (n+3) position. Hence again no cross-linked product is observed for the shorter linker, while two additional bands are observed for the longer linker.

We assumed that the guanine opposite to the reactive group forms weak hydrogen bonds with modified cytosine of the donor strand as well as with amide group of the reactive module that inhibits any flexibility of the reactive module. Surprisingly, the reaction with a template that contains a single mismatch each opposite the reactive group and the G in the (n+3) position reveals two bands in case of the shorter linker, but no band with the longer linker. It is not clear,

how 2 cross-linked products can be obtained. Next we decreased the duplex stability and hence increased the flexibility around the reactive group by placing A-residues opposite the A-residues (and G) of the reactive sequence in position (n-2), (n-1), (n+1) and (n+2) (ODN_4.34 – ODN_4.36). When the target guanine is placed opposite the reactive group (ODN_4.34) or when G is in the (n+3) position (ODN_4.35) a single cross-linked product is observed for both linker size probably because the shorter linker can now reach through the loop structure for reaction. However, also in the strand that contains two G, one opposite the reactive group and one in the (n+3) position (ODN_4.36) only one cross-linked product is observed. We assumed that due to sterical reasons, the G that is placed in the (n+3) position relative to the reactive group is a preferential target for modification while (n+0) does not seem to be accessible.

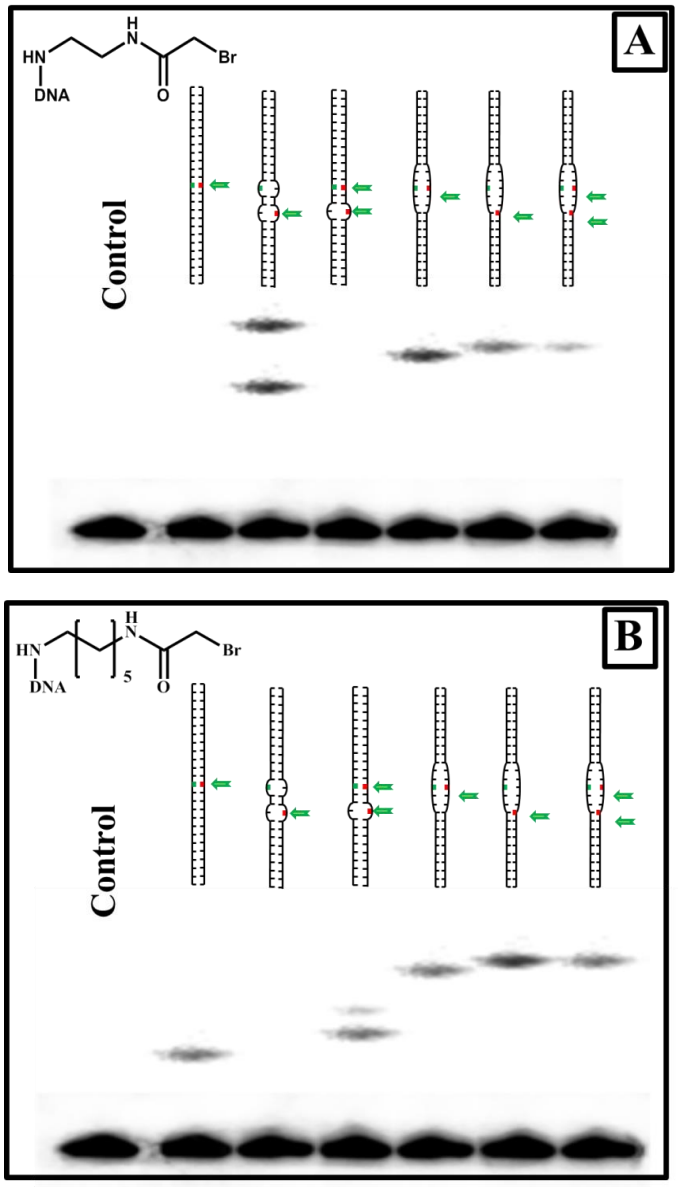


Figure 4.19. 12% Denaturing PAGE analysis of cross-link formation with locked constructs. SRS_4.11 containing either **A**: a diaminoethane (SRS_4.12) or **B**: a diaminohexane (SRS_4.13) linker was reacted with target sequences ODN_4.31 – ODN_4.36.

4.11 Stability of N7-guanine alkylated DNA primer

The alkylation of the N7 position of guanine creates a product that can undergo depurination followed by DNA degradation. We reasoned that although the N^7-alkyl-guanine modified nucleoside would likely be too unstable for long term applications, it might be possible to use this modification in biophysical measurements with fast kinetic parameters. Thus, elucidation of the influence of fluorophore incorporation on DNA stability is required to identify the limits of its application. Preliminary studies using different alkyl-guanosine derivatives revealed that these molecules are stable for weeks and would be suitable for G alkylation (**Table 4.1**). Specifically electron-withdrawing substitution of the alkane chain is a suitable candidate for cross-link formation with a stability profile that might allow its post-synthetical modification with different labels. We therefore tested the ability of guanine derivatives to mediate site-specific cross-link formation and their stability after labeling.

As a first test we alkylated a DNA oligonucleotide template (ODN_4.1) that contains only a single guanine with 2-bromoacetate and then analyzed the stability of the modified DNA template against depurination and strand degradation. After incubation with 2-bromoacetate for 24 hours corresponding to the incubation time required for G alkylation with the modular system, small organic compounds were removed by size-exclusion chromatography and hence the alkylation reaction was stopped. The first eluted fraction contained both initial and alkylated oligonucleotide. After this, the oligonucleotide containing mixture was placed at RT in 10 mM TEAA buffer at pH = 7.4. Aliquots were taken after every 24 hours and without any additional treatment directly measured by MALDI-TOF to confirm their identity and purity. Modified and nonmodified DNA is distinguished by minor changes in electronic structures, so we assume that the rate of ionization will be similar for both types. An additional advantage of this method is that usually only single charged ions are observed, hence peaks with the mass of the DNA are observed. As unmodified DNA is stable toward hydrolysis in neutral solution in contrast to modified DNA the signal of unmodified DNA can be used, to a certain extend, as internal standard for calibration. Accordingly, the stability of the alkylated DNA was quantified by comparison the intensity of its peak with the one of unmodified DNA ($M_{observed}$ = 6115.0 Da, M_{calc} = 6116.0 Da). The peak of modified DNA ($M_{observed}$ = 6174.0 Da, M_{calc} = 6174.3 Da) remains intensive without any changes during four days (**Fig. 4.20**). The intensity of the modified peak was equal to the one of the non-modified peak at the first day. This intensity we used as a starting point and assigned it to a value of 100% in stability of modified peak. In the next measurements we observed decreasing intensity of modified peak during four days according to normalized non-modified DNA strand (**Fig. 4.24**)

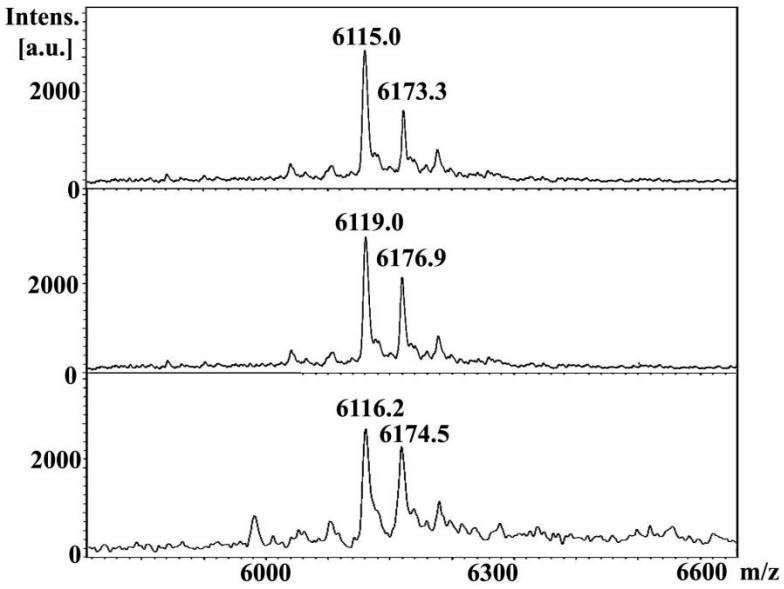


Figure 4.20. MALDI-TOF MS analysis stability of N7-acetate guanine modified ODN_4.1 at pH = 7.4 after 24, 48 and 96 hours.

In a next step we examined stability of N7-alkylated guanine in the cross-linked construct. As a reactive sequence we used 15-mer recognition sequences that should provide 15% of cross-link products after annealing (according to previous experiment, see **Fig. 4.15**) and introduced a cystamine cleavable linker as well as a 2-bromoacetate reactive group according to the previously describe procedure (see experimental part) resulting in reactive sequence SRS_4.10.

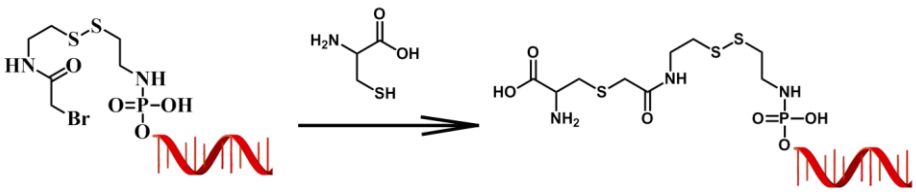
The reaction was started by annealing SRS_4.4 with the complementary radiolabeled oligonucleotide ODN_4.8 that contains G in the (n+1) position to the reactive group. The reaction was incubated for 24 hours in 1 M NaCl in 0.5 M MOPS buffer pH = 7.4 at RT. We used the same incubation time as in previous experiments. After incubation, the sample was divided into four parts and one of the parts was used as a starting point (0 days in the **Fig. 4.21, lane 2**) for the evaluation of DNA stability.

Second part we incubated at the same buffer pH = 7.4 and aliquots were taken after 0, 1, 2 and 4 days. All aliquots were stored at -20°C before gel analysis. After 12% denaturing PAGE we observed that the intensity of the band for the cross-linked product slowly increased during incubation time due to presence of 2-bromoacetate reactive group. To demonstrate that the upper band belongs to the cross-linked product and contains a disulfide cleavable linker we treated the sample after 4 days of incubation with TCEP that results in a drop in intensity of the upper band. Inactivation of the reactive group due to nucleophilic exchange of the bromine group against a hydroxyl-group is highly pH dependent and at pH = 7.4 finished after 7 days[32]. However, from this data is not possible to determine the stability of the alkylated compound, because decay and

formation of the compound take place in parallel as long as reactive group is still present in solution.

There are two methods to stop the reaction. As alkylation of guanine at position N7 is only efficient at neutral to basic pH, the third part of stock solution was acidified to pH = 5.6 after 24 hours to stop of the formation of new product (lane 7 – 10 on **Fig. 4.21**). Samples of the reaction at pH = 5.6 were taken every 24 hour. After gel analysis we observed almost a similar level of intensity of upper band for the first two days. Only after 48 hours decay of alkylated DNA strand was detected.

In the last experiment, we tried to inactivate the reaction of guanine modification by addition of excess of cysteine. According to reference[33] we assumed that unreacted 2-bromoacetate group will fast and efficiently modify the free thiol group from the cysteine moiety. Indeed, after PAGE analysis we did not observed increasing of band intensity, however, fast decay of cross-linked product was observed (**Fig. 4.22**). This can be explained by disulfide exchange of cysteine with the cleavable linker leading to cleavage of the disulfide linked duplex (**Scheme 4.6**).



Scheme 4.6. Quenching of reactive sequences by cysteine.

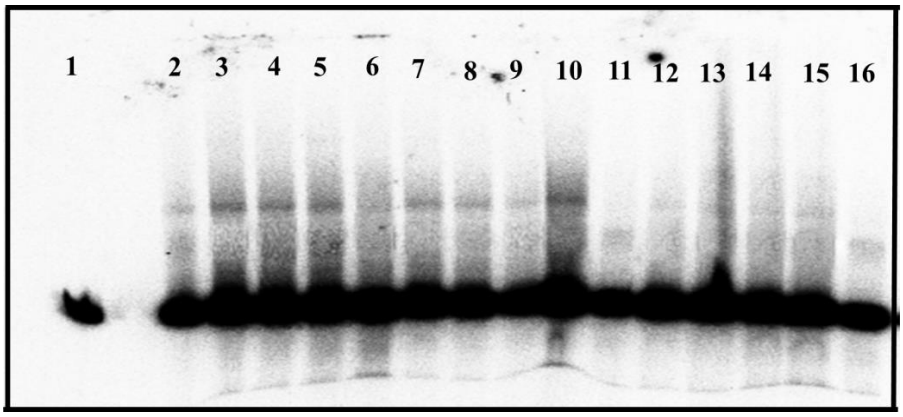


Figure 4.21. **A**: 12% denaturing PAGE analysis of stability of cross-linked product. Lane 1 – ODN_4.4; lane 2 – cross-linked product after incubation of SRS_4.4 with ODN_4.8 for 24 hours. This sample was used as a starting point (0 day, 100% of ICL peak stability) for the following three experiments, lanes 3 – 6: Reaction of guanine alkylation without quenching after

1 day, 2 days, 4 days, and TCEP cleavage; lanes 7 – 11: Reaction of guanine alkylation was quenched by decreasing pH = 5.6 (after 1 day, 2 days, 3 days, 4 days, and TCEP cleavage); lanes 12 – 16: Reaction of guanine alkylation was quenched by adding of cysteine (after 1 day, 2 days, 3 days, 4 days, and TCEP cleavage).

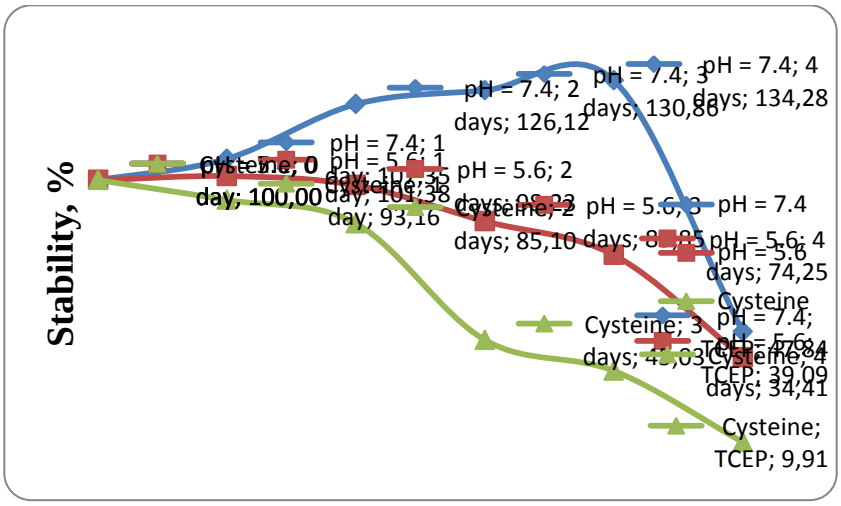


Figure 4.22. Evaluation of stability of cross-linked product in PAGE analysis. Blue curve – reaction without quenching; green curve – cysteine quenching; red curve – pH quenching.

We then studied the influence of fluorophore attachment on the DNA strand stability. In comparison with previously reported mBBr modification, in this experiment we used 17-mer recognition sequence SRS_4.5 that forms a stable duplex with complementary target strand ODN_4.19 ($T_{melting}$ = 53°C). According to our approximation of duplex stability versus cross-link efficiency a yield up to 10% - 15% is expected. Reactive sequence SRS_4.6 includes a cystamine cleavable site and a 2-bromoacetate reactive group, and was synthetized according to the previously described protocol (see experimental part).

The reactions were carried out at 1.3 μM DNA concentration in MOPS buffer, pH = 7.4, containing 1 M NaCl and incubated for 24 hours. After incubation, reaction was treated with TCEP for two hours under nitrogen and 5 eq. of excess of mBBr were added. After three hours the reaction was purified by Nap-5 size-exclusion column and lyophilized. Lyophilized sample was redissolved in 0.1 M MOPS buffer, pH = 7.4 in presence of 1 M NaCl and we assigned it as starting point (0 day). Aliquots were taken out after 1, 2 and 4 days and quantified by fluorescence image of reverse-phase HPLC chromatogram. Peak at 15.2 min corresponds to the mBBr-labeled DNA template and slowly degraded to give a product with retention time of 18.2 min (**Fig. 4.23 B**). This spectrum also contains the peak of labeled SRS_4.6 and low amounts of mBBr adducts, but the yield varies. Overall, the decomposition was relatively high (> 90% in 4

days), possibly due to a fluorophore degradation in addition to DNA cleavage. In this experiment we assumed, that intensity of mBBr-labeled SRS_4.6 peak and mBBr peak is constant for 4 days (**Fig. 4.24**). For all three types of stability experiment, the half-life of alkylated DNA products differed from 60 to 72 hours. It is not yet clear whether this difference is caused by difference in electronic structure of side chain in guanine (from simple 2-bromoacetate side chain in first trial to bulky mBBr substitution in the last) or was due to the experimental detection method used.

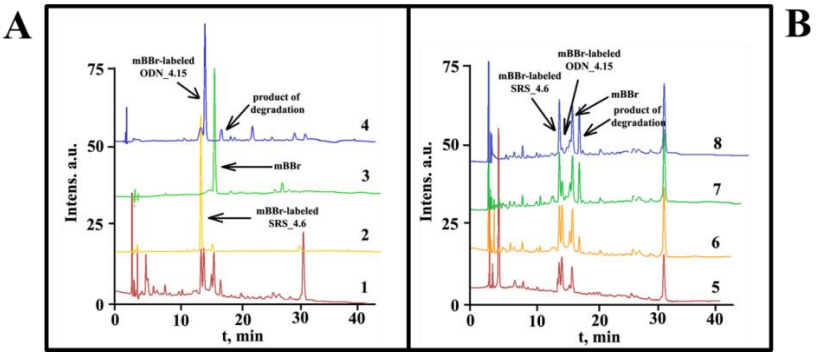


Figure 4.23. Reverse-phase HPLC evaluation of the stability of mBBr-labelled oligonucleotide (λ = 390 nm). **A**: RP-HPLC chromatogram of reaction mixture: **1**: mixture of oligonucleotides after mBBr labelling; **2**: mBBr-labeled SRS_4.6; **3**: pure mBBr; **4**: mBBr-labelled ODN_4.4. **B**: Stability of mBBr-labelled oligonucleotides: **5**: 0 day; **6**: 1 day; **7**: 2 days; **8**: 4 days.

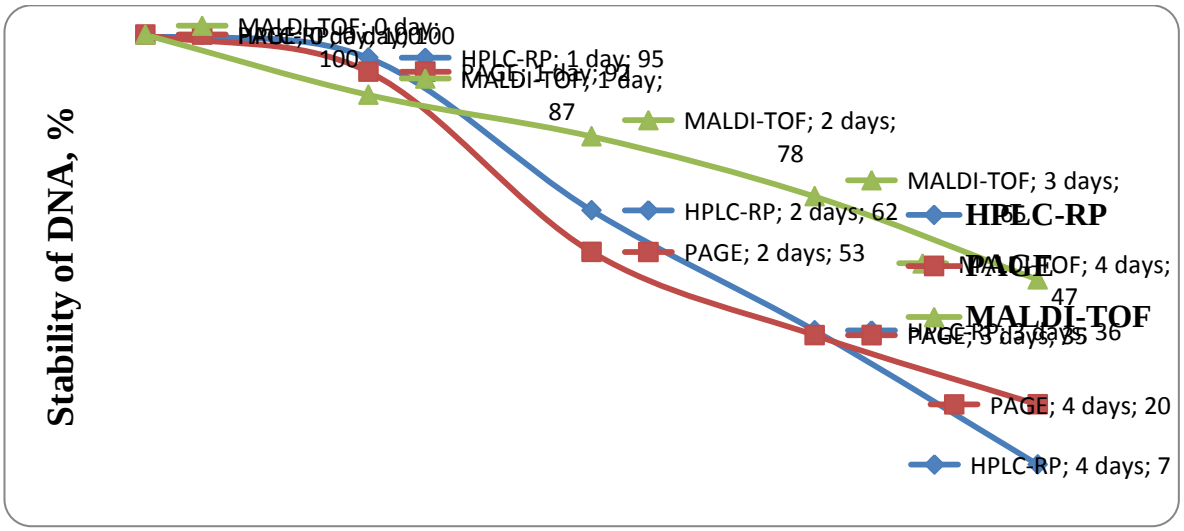


Figure 4.24. Stability of N7-alkylated guanine DNA primer analyzed with different methods: diamond – RP-HPLC analysis of stability of mBBr-oligonucleotide adduct (**Fig. 4.23**); square – PAGE analysis of stability of cross-linked product (**Fig. 4.22**); triangle – MALDI-TOF MS analysis of nonspecific modified oligonucleotide (**Fig. 4.20**).

4.12 Conclusions

The collected data shows that the modular system can be used for site-specific modifications of long DNA or RNA, respectively. The reaction system consists of three building blocks that can be synthesized independently manifesting the proposed simple controllable way for nucleic acid modifications. Alkylation of guanine was successfully applied to introduce a functional group, i.e. a thiol that can be used for labeling purposes.

Our findings also demonstrate how the specificity of modifications can be altered by changes of the template or the linker architecture. Additionally, we identified the most reactive position in the target strand and the distance-dependent modification of nucleic bases relative to the position of the reactive group.

The yield of modification and labeling directly depends on the stability of the duplex formed between reactive and target sequences. In general, the limited stability of N7 alkylated DNA hampers its use in long term biophysical applications. Nevertheless, labeling of guanine with fluorophore using the presented system can be used for applications in which a fast labeling and detection is required. Taken together, these results expand the range of DNA-template modifications.

4.13 References

[1] J. H. P. Chan, S. Lim, W. S. F. Wong, *Clin. Exp. Pharmacol. P.* **2006**, *33*, 533-540.
[2] K. M. Vasquez, T. G. Wensel, M. E. Hogan, J. H. Wilson, *Biochemistry* **1996**, *35*, 10712-10719.
[3] J. Tilsner, C. Flors, *ChemBioChem* **2011**, *12*, 1007-1009.
[4] R. Ekkebus, S. I. van Kasteren, Y. Kulathu, A. Scholten, I. Berlin, P. P. Geurink, A. de Jong, S. Goerdayal, J. Neefjes, A. J. R. Heck, D. Komander, H. Ovaa, *J. Am. Chem. Soc.* **2013**, *135*, 2867-2870.
[5] G. Sengle, A. Eisenführ, P. S. Arora, J. S. Nowick, M. Famulok, *Chemistry & Biology* **2001**, *8*, 459-473.
[6] J. Hall, D. Hüsken, R. Häner, *Nucleic Acids Res.* **1996**, *24*, 3522-3526.
[7] J. C. Joyner, J. A. Cowan, *J. Am. Chem. Soc.* **2011**, *133*, 9912-9922.
[8] D. M. Kolpashchikov, *Chem. Rev.* **2010**, *110*, 4709-4723.
[9] A. Sancar, L. A. Lindsey-Boltz, K. Ünsal-Kaçmaz, S. Linn, *Annu. Rev. Biochem* **2004**, *73*, 39-85.
[10] A. Pullman, B. Pullman, *Q. Rev. Biophys.* **1981**, *14*, 289-380.
[11] B. Reiner, S. Zamenhof, *J. Biol. Chem.* **1957**, *228*, 475-486.
[12] M. Zewail-Foote, L. H. Hurley, *J. Am. Chem. Soc.* **2001**, *123*, 6485-6495.
[13] K. S. Gates, T. Nooner, S. Dutta, *Chem. Res. Toxicol.* **2004**, *17*, 839-856.
[14] J. A. Zoltewicz, D. F. Clark, T. W. Sharpless, G. Grahe, *J. Am. Chem. Soc.* **1970**, *92*, 1741-1750.
[15] R. C. Moschel, W. R. Hudgins, A. Dipple, *J. Org. Chem.* **1984**, *49*, 363-372.
[16] M. R. Osborne, D. H. Phillips, *Chem. Res. Toxicol.* **2000**, *13*, 257-261.
[17] N. Fedtke, J. A. Boucheron, V. E. Walker, J. A. Swenberg, *Carcinogenesis* **1990**, *11*, 1287-1292.
[18] G. L. Foureman, D. J. Reed, *Biochemistry* **1987**, *26*, 2028-2033.
[19] P. Brookes, P. D. Lawley, *Biochem. J.* **1961**, *80*, 496-503.
[20] P. B. Inskeep, N. Koga, J. L. Cmarik, F. P. Guengerich, *Cancer Res.* **1986**, *46*, 2839-2844.
[21] K. Hemminki, *Chem. Biol. Interact.* **1984**, *48*, 249-260.
[22] G. N. Grimm, A. S. Boutorine, C. Helene, *Nucleos. Nucleot. Nucl.* **2000**, *19*, 1943-1965.
[23] D. G. Knorre, P. V. Alekseyev, Y. V. Gerassimova, V. N. Silnikov, G. A. Maksakova, T. S. Godovikova, *Nucleos. Nucleot.* **1998**, *17*, 397-410.
[24] C. R. Calladine, H. R. Drew, B. F. Luisi, A. A. Travers, *Understanding DNA: the molecules and how is it work*, third addition ed., Elsevier Academic Press, **2004**.

[25] M. Rueping, *Angew. Chem. Int. Ed.* **2006**, *45*, 1838-1840.
[26] B. Zhang, T. R. Cech, *Nature* **1997**, *390*, 96-100.
[27] P. J. Unrau, D. P. Bartel, *Nature* **1998**, *395*, 260-263.
[28] R. F. Gesteland, T. R. Cech, J. F. Atkins, *The RNA World*, 2nd ed., Cold Spring Harbor Press, Cold Spring Harbor, **1999**.
[29] H. Moser, P. Dervan, *Science* **1987**, *238*, 645-650.
[30] K. M. Vasquez, P. M. Glazer, *Q. Rev. Biophys.* **2002**, *35*, 89-107.
[31] L. Christensen, R. Fitzpatrick, B. Gildea, K. H. Petersen, H. F. Hansen, T. Koch, M. Egholm, O. Buchardt, P. E. Nielsen, J. Coull, R. H. Berg, *J. Pept. Sci.* **1995**, *1*, 175-183.
[32] M. J. Taylor, P. B. Dervan, *Bioconjugate Chem.* **1997**, *8*, 354-364.
[33] M. Trmčić, D. R. W. Hodgson, *Beilstein J. Org. Chem.* **2010**, *6*, 732-741.

CHAPTER 5

Fast and effective alkylation of the N3 position in thymidine with electrophiles

5.1 Introduction

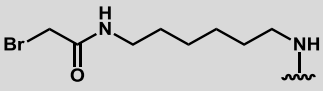
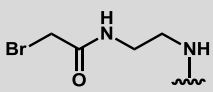
Pyrimidine is an aromatic system with irregular distribution of electron density. This is caused by the following two factors, namely, induction and mesomeric effects of nitrogens[1]. It leads to the formation of pronounced partially positive charge on the carbon atoms in positions 2, 4, and 6. Due to this fact, the aromatic pyrimidine ring is stable towards electrophilic reagents to an even greater extent than pyridine. The only positions available for modification by electrophilic substitution are the nitrogen in the N3 position, the 5,6-double bond and the position C5.

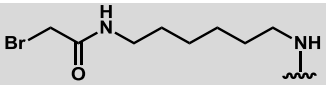
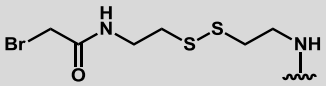
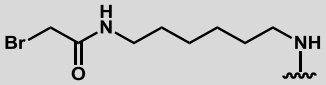
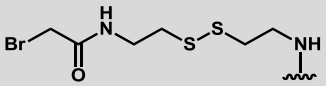
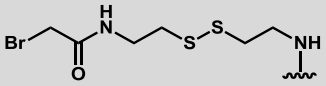
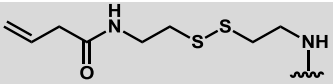
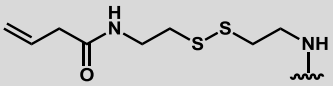
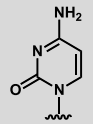
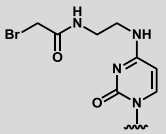
Nevertheless, conditions that are different from physiological, can direct or activate additional reactive centers in the thymine heterocycle. Position N3 in thymine forms three σ-bonds with two carbon atoms and one hydrogen, respectively. Of great interest in the N ─ H bond, that has a pK_α value of 9[2] can be deprotonated at slight alkaline conditions forming a free electron pair on the N3 nitrogen. This position can be attacked by different electrophiles, e.g. halogen-derivatives[3,4] or acrylic acid[5].

In this chapter we explore an approach for site-specific modification of the activated N3 position in thymidine within an oligonucleotide.

Sequences:

*Complementary regions are underlined

Name	Reactive group	Sequence
ODN_5.1	─	5’- CAC AC**T T**A**T** AA**G** AAA AAA AA -3’
SRS_5.1	PO_3H	5’-X-<u>CAG AGA AGG G</u>-3’
SRS_5.2		5’-X-<u>CAG AGA AGG G</u>-3’
ODN_5.2	─	5’-<u>CCC TTC TC</u>**T TTT T**-3’
ODN_5.3	─	5’-<u>CCC TTC TC</u>**A AAA A**-3’
ODN_5.4	─	5’-<u>CCC TTC TC</u>**C CCC C**-3’
SRS_5.3	PO_3H	5’-X-<u>TTT GGT GGT G</u> -3’
SRS_5.4		5’-X-<u>TTT GGT GGT G</u> -3’

SRS_5.5		5’-X-TTT GGT GGT G -3’
SRS_5.6		5’-X-TTT GGT GGT G -3’
SRS_5.7	PO_3H	5’-X-TTT TTTTTT TGG TGG TG -3’
SRS_5.8		5’-X-TTT TTTTTT TGG TGG TG -3’
SRS_5.9		5’-X-TTT TTTTTT TGG TGG TG -3’
ODN_5.5	–	5’-CAC CAC CAA A**T**A AAA AAA AA-3’
ODN_5.6	–	5’-CAC CAC CAA AA**T** AAA AAA AA-3’
ODN_5.7	–	5’-CAC CAC CAA AAA **T**AA AAA AA-3’
ODN_5.8	–	5’-CAC CAC CAA AAA A**T**A AAA AA-3’
ODN_5.9	–	5’-CAC CAC CAA AAA AA**T** AAA AA-3’
ODN_5.10	–	5’-CAC CAC CAA AAA AAA **T**AA AA-3’
ODN_5.11	–	5’-CAC CAC CAA AAA AAA A**T**A AA-3’
ODN_5.12	–	5’-CAC CAC CAA AAA AAA AA**T** AA-3’
ODN_5.13	–	5’-CAC CAC CAA AAA AAA AAA **T**A-3’
ODN_5.14	–	5’-CAC CAC CAA AAA AAA AAA A**T**-3’
SRS_5.10	PO_3H	5’- AT GCG GCA GGC TGC A-3’
SRS_5.11		5’- AT GCG GCA GGC TGC A-3’
ODN_5.15	–	5’-TGC AGC CTG CCG CA**T** **T**AC TCG TGG T– 3’
SRS_5.12		5’-X-TTT GGT GGT G -3’
SRS_5.13		5’-X-TTT TTTTTT TGG TGG TG -3’
SRS_5.14		5’- GGT ATG GAT TAA CAA TTA GTG TGG T -3’
SRS_5.15		5’- GGT ATG GAT TAA CAA TTA GTG TGG T -3’
ODN_5.16	–	5’-ACC ACA CTA AAA **T**AA AAT CCA TAC C – 3’

5.2 Theoretical background of thymidine alkylation

The properties of isolated pyrimidine and purine bases are primarily determined by the electronic structure and hence the spatial distribution of electron density at each reactive atom of the base[6]. Comparison of the electronic structure of nucleobases leads to the identification of reactive sites that are selective for different types of reactions. Such reactivity prediction can be performed based on two different theories. The first theory considers the shape and free energy of the molecular orbitals in the conjugated system[7]. The nitrogen atoms in purine and pyrimidine bases exist in pyridine and pyrrole forms. The pyrrole nitrogens are sp^2 hybridized and hence form three σ-bonds at 120° angles with carbons or hydrogens with the orbitals lying in the plane of the aromatic ring. Additionally, two electrons of the p-orbital of the nitrogen overlap with a p-orbital of one of the carbons. This orbital lies perpendicular to the plane of the heterocycle. Nitrogen atoms of this type are available for electrophilic attack. However, the attack will only be efficient if the reagent is also located in direction perpendicular to the plane of the heterocyclic base.

The second theory is based on the acid-base properties of the heterocyclic base[8] (**Table 5.1**). Accordingly, the pH value of the reaction solution can be used to modulate reactivity of nucleobases. Nucleobases exist in their charge neutral form under physiological conditions. However, protonation and deprotonation of each base type is possible under the appropriate aqueous condition resulting in the charged form. Based on the pK_α values of all four nucleobases only the N3 positions of uracil and thymine and the N1 position of guanine can be deprotonated in alkaline aqueous solution and yield the anionic base form.

Table 5.1. pK_α values of nucleotides and their components[2].

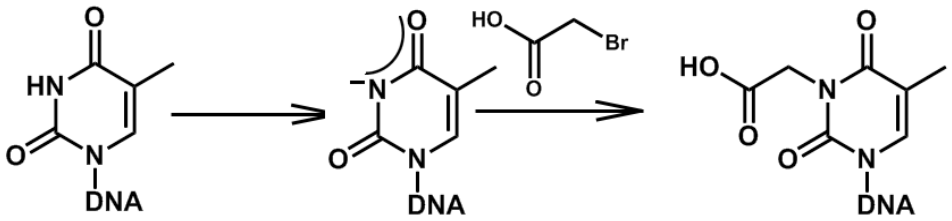
Compound	pK_α

	base	pentose	phosphate group	
			pK_1	pK_2
Cytosine (N3, N4)	4.4[9]; 12.2			
Deoxycytidine (N3)	4.25			
		12.24		
Cytidine 5'-phosphate (N3)	4.5		0.8	6.3
Cytidine 3'-phosphate (N3)	4.16		0.8	6.04
Uracil (N3)	9.2[8]			
Deoxyuridine (N3)	9.3			
		12.59		
Uridine 5'-phosphate (N3)	9.5		1.0	6.4
Uridine3'-phosphates (N3)	9.96		1.0	5.9
Thymine (N3)	9.9[10]			
Deoxythymidine (N3)	9.8	12.85		
Deoxythymidine 5'-phosphate (N3)	10.0		1.6	6.5
Adenine (N1)	4.1[10]			
Deoxyadenosine (N1)	3.8			
		12.35		
Adenosine 5'-phosphate (N1)	3.74		0.9	6.05
Adenosine 3'-phosphates (N1)	3.7		0.9	6.1
Guanine (N1, N7)	3.2; 9.5[11]			
Deoxyguanosine (N1, N7)	2.4; 9.33			
		12.3		
Guanosine 5'-phosphate (N1, N7)	2.9; 9.6		0.7	6.3
Guanosine3'-phosphates (N1, N7)	2.4; 9.8		0.7	6.0

5.3 Alkylations of mononucleotides with bromoacetate

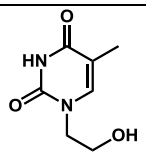
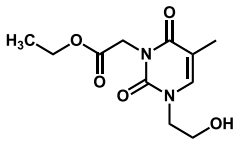
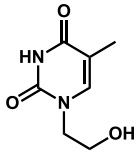
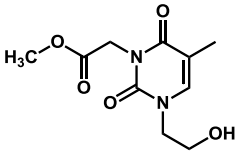
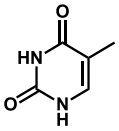
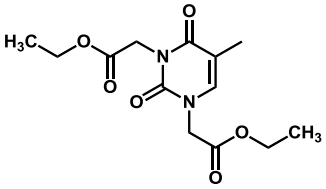
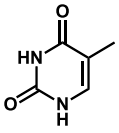
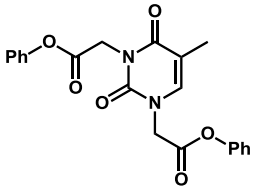
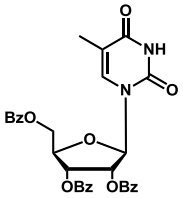
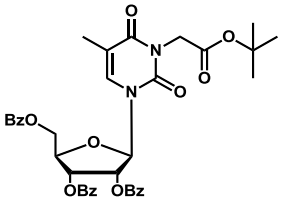
The most popular strategy for alkylation of thymidine at the N3 position involves treatment of the monomeric nucleobase with 2-bromoacetate derivatives under strong alkaline conditions

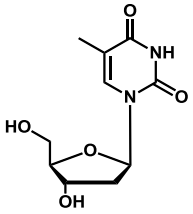
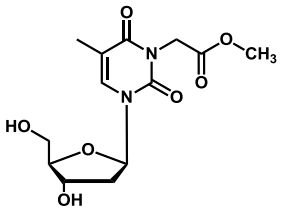
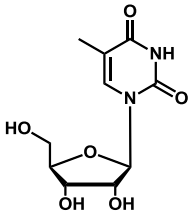
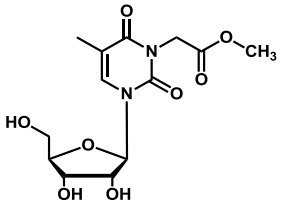
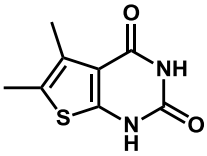
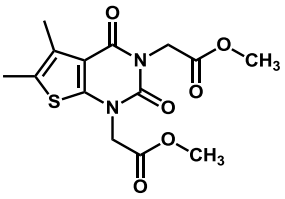
in different organic solvents (**Table 5.2**). For the incorporation of modified mononucleotides into DNA primers solid phase synthesis is used. To our knowledge, only few publications up to now describe modification of the N3-position of thymidine within an oligonucleotide[12]. Considering the pK_α value of 9.9 of thymine N3 position, 2-bromoacetate should react with the N3 atom under conditions that are still sufficient to retain the secondary structure of DNA (**Scheme 5.1**).



Scheme 5.1. Alkylation of thymidine with 2-bromoacetic acid under alkaline conditions.

Table 5.2. Different reactions for N3 thymidine alkylation.

Entry	Thymine derivative	Conditions[a]	Product	Ref
1		K_2CO_3, DMF, 3days		3
2		NaH, DMF, 15 min, 80°C		13
3		NaH, DMF, 3 min, 120°C		4
4		K_2CO_3, NaH, DMF, 24 hours		14
5		K_2CO_3, DMF, 24 hours		15

6		K_2CO_3, MeOH, 6 hours, reflux		16
7		K_2CO_3, Me_2CO, 6 hours, reflux		17
8		NaOH, TEBAC, $CH_2Cl_2 – H_2O$		9

a) Respective bromine-compound was used as a reactant in all reactions.

To evaluate the reactivity of all four deoxyribonucleotides under modification conditions separate treatment with 2-bromoacetate at pH = 9.3 was performed for 12 h at 37°C (**Fig. 5.1 A**). As expected from the chemical structure of the nucleosides only thymidine and guanosine showed formation of covalent adducts at alkaline pH.

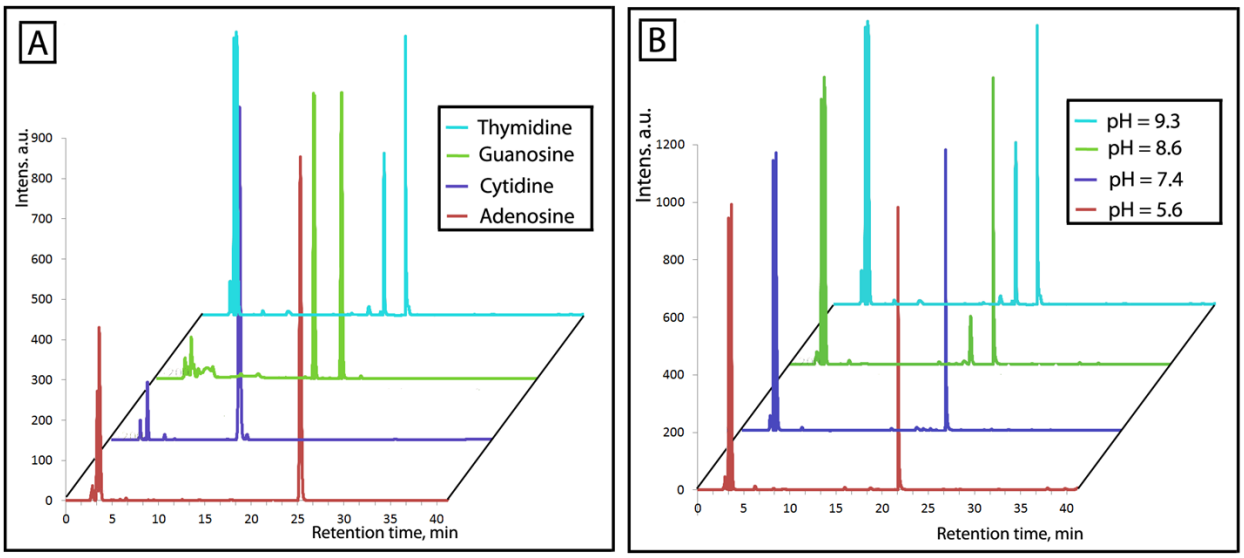


Figure 5.1. Reverse-phase HPLC chromatograms of **A:** alkylations of four deoxynucleosides under basic conditions represent formation of new products with G and T residues. **B:** alkylation of thymidine with 2-bromoacetic acid in different buffers demonstrated

its activity only at alkaline conditions; pH values in the legend demonstrate pH value of solution after addition of both reactive compounds. Stock solution of 2-bromoacetic acid was adjusted to the desired pH with 0.1 M NaOH and immediately added to the reaction.

To evaluate the pH dependence of the reaction with thymidine different buffers were used. (**Fig. 5.1 B**). We expected the highest alkylation yields at a pH equal to the pK_{α} of the N3 position. Indeed we observed 37% of thymidine alkylation and less at lower pH values until the reaction virtually stopped at neutral pH.

To verify pH-dependent alkylation LC-MS analysis was conducted. Under acidic conditions a single peak at $\lambda = 260$ nm was observed with an elution time of 16.8 min and a mass of 265.1 Da corresponding to the $[M+H+Na]^+$ signal ($M_{calc} = 265.1$ Da) (**Fig. 5.2 A and C**). Co-injection of thymidine demonstrated identical ESI-MS and UV spectrum with $\lambda_{max} = 262$ nm. In contrast, the reaction at basic pH gave an additional peak that eluted at 21.2 min (**Fig. 5.2 B**). The tandem mass spectrum gives a molecular mass of this peak of m/z 323.1 Da in the positive mode and of m/z 298.9 Da in the negative mode respectively ($[M+H+Na]^+_{calc.}$ 323.06 Da; $[M-H]^-_{calc.}$ 299.2 Da) (**Fig. 5.2 C**). The formation of easily ionized compounds in the negative mode suggests addition of a carboxylic group to the thymidine molecule.

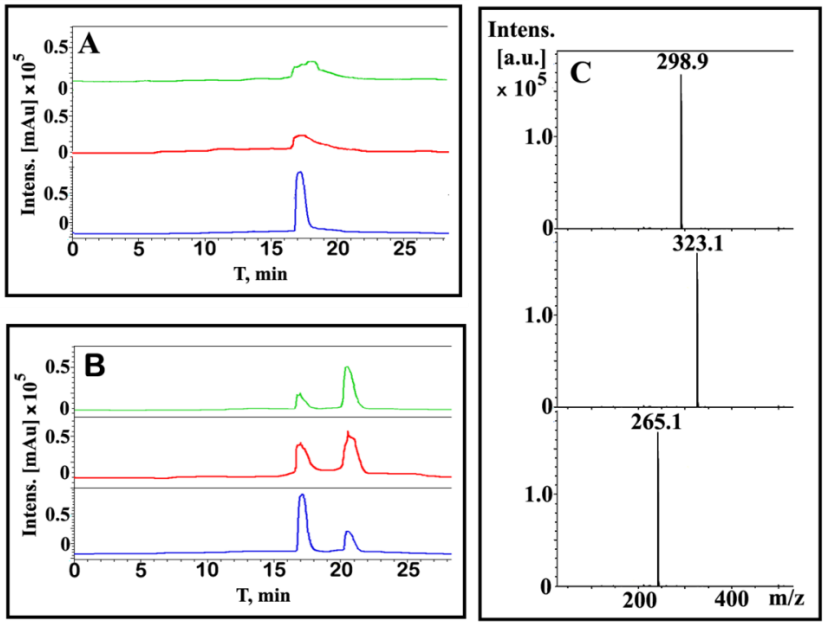


Figure 5.2. Reverse-phase HPLC chromatogram with different detectors ($\lambda = 260$ nm – blue line, ESI positive mode – red line, ESI negative mode –green line) of: **A:** reaction of thymidine with bromoacetate at pH=5.6. **B**: reaction of thymidine with bromoacetate at pH = 9.3 **C**: Deconvoluted ESI-MS spectra of 1) Thymidine eluting at 16.8 min (calc. mass 265.09 Da), 2)

modified thymidine eluting at 21.2 min in positive mode ($[M+H+Na]^+$calc. 323.06 Da), 3) modified thymidine eluting at 21.2 min in negative mode ($[M-H]^-$calc. 299.22 Da)

One of the expected side reactions during thymidine alkylation is the inactivation of the reactive group due to nucleophilic substitution of the bromine atom with a hydroxyl group under basic conditions[18]. In a control experiment we hydrolyzed 2-bromo-acetate for 12 hours at pH = 9.3 followed by reaction with thymidine. The HPLC chromatogram did not show formation of a peak for alkylated thymidine. One of the possible ways to avoid inactivation of the reactive group is to replace it with a less reactive one, i.e. to replace the bromine by a chlorine atom.

According to theoretical calculations[19] a chlorine atom generates a similar positive charge on the C2 carbon atom of the acetate-moiety but is much more stable to hydrolysis. These properties should make it a suitable alternative for the bromo-reactive group. Indeed, LC-MS analysis of chloroacetate-treated thymidine with other HPLC column, displayed a clearly visible peak at t = 14.3 min responsible for the desired product. Screening this peak for the expected mass of an adduct of acetate with thymidine, provided a $[M+H]^+$ signal of m/z 299 Da. Unfortunately, this approach demonstrated lower yields, as the highest observed yield was around 25% in comparison with 37% from the previous experiment.

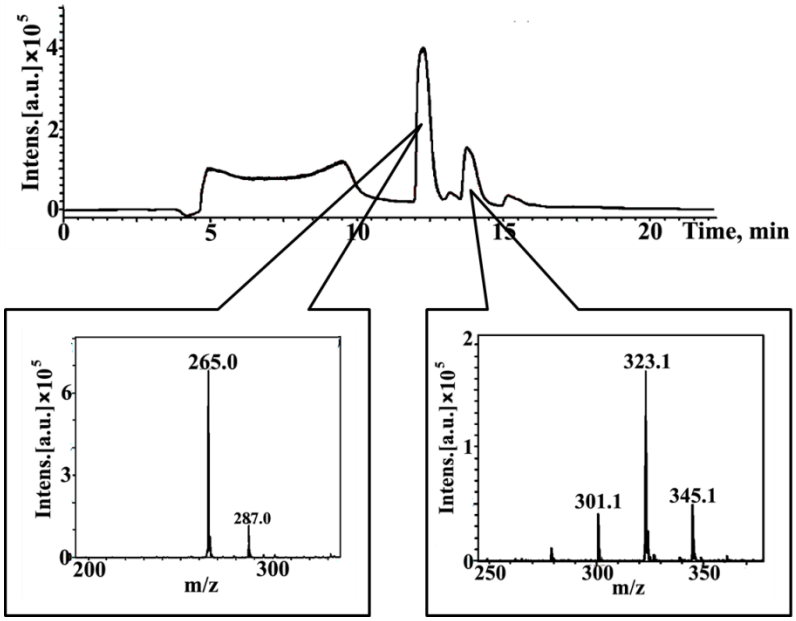


Figure 5.3. LC-MS analysis of N3 thymidine alkylation with 2-chloroacetate at 0.1 M $NaHCO_3$, pH = 9.3. Thymidine eluting at 13.5 min (calc. mass 265.09 Da), 2) modified thymidine eluting at 14.3 min in positive mode ($[M+H]^+_{calc.}$ = 300.16 Da) ($[M+H+Na]^+_{calc.}$ = 323.06 Da). Presence of additional peaks in both spectrum with molecular mass at 287.0 Da and 345.1 Da we correspond to attachment two sodium atoms to two carbonyl group in the heterocyclic core.

To increase the yield we tried next to use a more reactive group, i.e. iodine. As hydrolysis of this group is also increased, we intended an *in situ* exchange of chlorine to iodine in the presence of an equimolar amount of sodium iodide. After reaction mixtures produced HPLC-chromatograms and ESI spectra similar to the ones reported with chloroacetate as a reactive group and a moderate yield increase (31% compared to 25%).

Another approach to increase the reactivity of chloroacetate is to incorporate an additional electron-withdrawing group in the conjugated system of the reactive centre. We tested 2,2-dichloroacetate for its ability to alkylate thymidine. Unfortunately, we did not observe any new product peaks, what can be explained by fast hydrolysis of 2,2-dichloroacetate in basic solution.

5.4 Reaction of 2-bromoacetate with short DNA oligonucleotides

In the next step of analysis we attempted the nonspecific modification of a DNA primer. For this, the sequence ODN_5.1 which contains one guanine and three thymidines for potential modification was treated with 2-bromoacetate or 2-chloroacetate, respectively. After reaction, small organic molecules were removed from reaction by size-exclusion chromatography and the oligonucleotide solutions were analyzed by MALDI-TOF MS without any further purification (**Fig. 5.4**). For both reactions, the MALDI-TOF demonstrated three major peaks with masses of 6116.8 Da ($M_{calc.}$ = 6118.3 Da) corresponding to the unmodified DNA primer, 6174.8 Da for the DNA primer with one acetate side chain ($M_{calc.}$ = 6175.9 Da) and 6232.8 Da for the primer with two acetate moieties ($M_{calc.}$ = 6234.1 Da). The precise sites of modification were not determined, but according to general reactivity the first alkylation site should be guanine – N7 followed by thymine – N3.

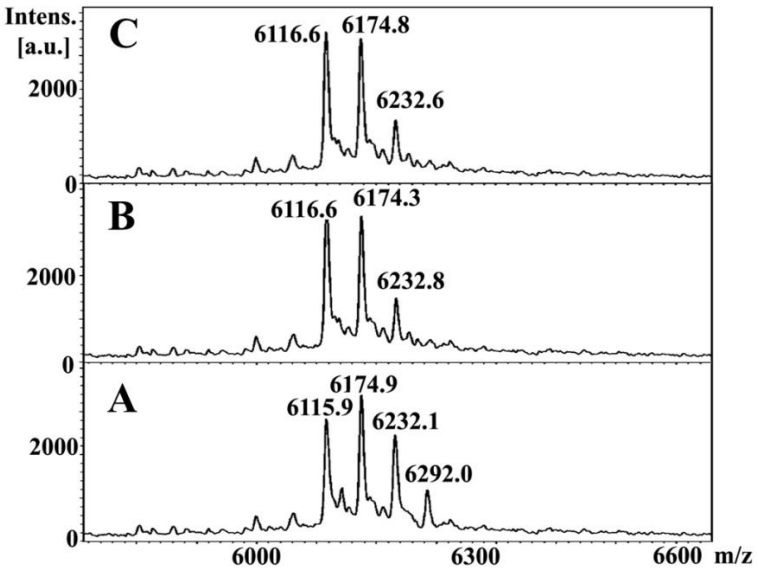


Figure 5.4. MALDI-TOF MS of ODN_5.1 that was non-specifically modified with: **A:** 2-bromoacetate; **B:** 2-cloroacetate; **C:** 2-chloroacetate in presence of NaI. Calculated mass of unmodified template is 6117.0 Da; Template modified with one acetate is 6175.8 Da; Template modified with two acetate moieties is 6232.1 Da; Template modified with three moieties is 6290.0 Da.

5.5 Synthesis of modular system for alkylation and evaluation of specificity for thymidine

It is well-known that duplex DNA is destabilized in basic solution as required for alkylation. Accordingly, we first evaluated the duplex stability of our system under reaction conditions. For this UV-melting experiment was performed using the unmodified 10-mer oligonucleotide primer SRS_5.1 and the complementary templates forming 8-mer (ODN_5.3 and ODN_5.4) and 9-mer helixes (ODN_5.2) (**Table 5.3**). The analysis of the melting profiles demonstrated a destabilization of duplex formation under basic conditions of 7 - 9°C. However, duplex stability is still sufficient for the planned experiments.

Table 5.3. Melting temperatures [in °C] of donor strand (SRS_5.1) with complementary sequences with different nucleotides in the overhang region at neutral and alkaline pH.

Sequence	T, °C, pH = 7.4	T, °C, pH = 9.3	ΔT, °C
ODN_5.2: 5'-CCC TTC TCT TTT T-3'	36	27	9
ODN_5.3: 5'-CCC TTC TCA AAA A-3'	34	27	7
ODN_5.4: 5'-CCC TTC TCC CCC C-3'	34	26	8

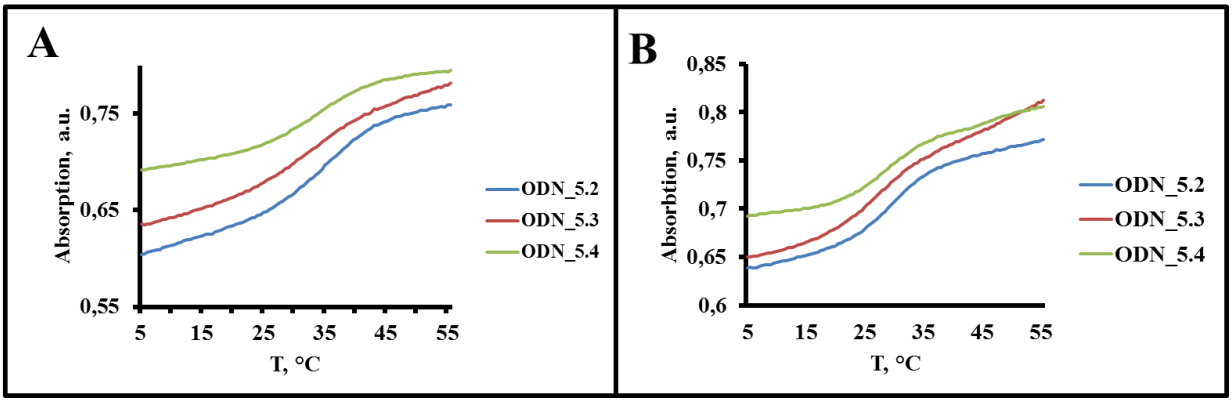
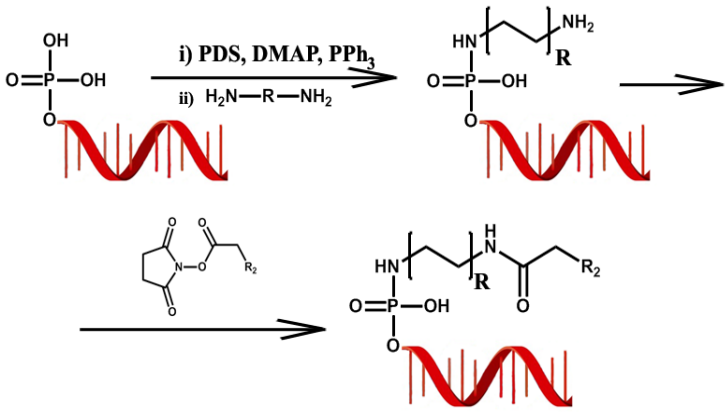


Figure 5.5. UV melting studies of donor strand SRS_5.1 with complementary sequences with different nucleotides in the overhang region: **A**: at neutral; **B**: alkaline pH. Spectrum were recorded at 260 nm, 2 mM duplex concentration, 100 mM sodium acetate buffer pH = 7.4 or sodium hydrogenarbonate buffer pH = 9.3 and 100 mM NaCl).



Scheme 5.2. Synthesis of reactive sequences by formation of the phosphoramidate bond between amino-containing molecule and guided sequence by Boutorine protocol[12] with following addition of NHS-ester of reactive group.

The activity of SRS_5.2 in cross-link formation was evaluated by reaction with different templates (ODN_5.2 – ODN_5.4) under different conditions and analyzed by SDS-PAGE **(Fig. 5.6**). To verify that A and C bases in the single stranded overhang of the template strand opposite the reactive group are also not modified under the conditions of the modular system, also the two templates ODN_5.3 and ODN_5.4 were tested, which contain five A of five C bases in the overhang, respectively. Indeed, under alkaline conditions and RT only the template ODN_5.2 with T residues results in the formation of cross-linked product with a yield of 62% (lane 2) while adenosine (ODN_5.3) (lane 7) and cytidine (ODN_5.4) (lane 12) did not show any alkylation and hence cross-link formation. As decreased incubation temperature should stabilize the duplex structure, the reactions were also performed at +4°C. However, apparently this decreased the reaction speed and consequently also T shows a decreased efficiency of cross-link product formation (yield is 18%) (lane 4). With A (ODN_5.3) and C (ODN_5.4) (lanes 8 and 14) again no cross-linking occurred.

When the reaction was performed at neutral pH for none of the three templates cross-link formation was observed, neither at RT nor at +4°C demonstrating the importance of alkaline conditions (**Fig. 5.6**). Hence in summary, reaction only takes place at thymine residues and not at adenine and cytosine at alkaline pH at RT. The occurrence of multiple bands (lane 2) indicates a

relatively low site-specificity. In addition, these results demonstrate two outcomes: 1) a reactive group consisting of a bromoacetate-derivative selectively forms covalently cross-linked product with thymidine residues but not with A and C and 2) site-specificity is rather low.

№	1		2	3	4	5	6	7	8	9	10	11	12	13	14	15
Nucleotide	T		T	T	T	T	A	A	A	A	A	C	C	C	C	C
SRS			+	+	+	+		+	+	+	+		+	+	+	+
Buffer	9		9	7	9	7	9	9	7	9	7	9	9	7	9	7
T, °C	20		20	20	4	4	20	20	20	4	4	20	20	20	4	4

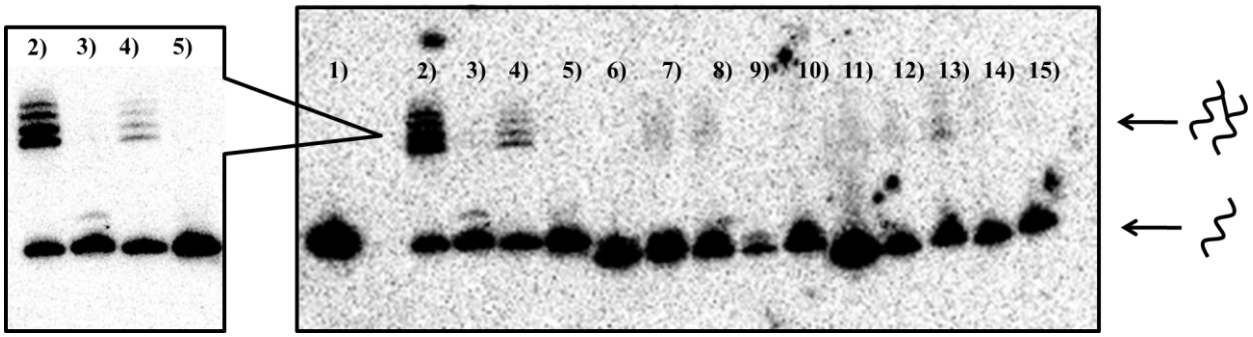


Figure 5.6. 18% Denaturing PAGE analysis of cross-link formation between 2-bromoacetate containing reactive sequence SRS_5.2 and appropriate template in 0.1 M $NaHCO_3$, pH = 9.3 or 0.1 M NaOAc, pH = 7.4 and 0.1 M NaCl for both buffers. Complementary templates contain pentanucleotide "T" (ODN_5.2) (lanes 1-5), "A" (ODN_5.3) (lanes 6-10) or "C" (ODN_5.4) (lanes 11-15) residues. See explanation of reaction conditions in the text.

5.6 Site-specificity of interstrand cross-link formation with modular system with non-cleavable and cleavable linkers

For the synthesis of the modular system the recognition sequence SRS_5.3 was modified at its 5'-phosphate group with a diamine according to Boutorine et al[12], i.e. with 1,2-diaminoethane (SRS_5.4), 1,6-diaminohexane (SRS_5.5) and cystamine (SRS_5.6) (**Scheme 5.2**). 2-bromoacetate was converted into the N-hydroxysuccinimide ester by reaction with N-hydroxysuccinimide and DCC and attached to the diamine-modified reactive sequence according to the previously published protocol[20]. After amide bond formation, the reactive DNA sequence was precipitated with ethanol and purified by RP-HPLC. In the last step the reactive sequence was characterized by MALDI-TOF MS and the desired peak of reactive sequence was identified (see Experimental part).

To analyse the site-specificity of cross-link formation, we investigated the reaction between recognition sequences with different linkers and complementary template sequences, which contain the target thymine at position (n+1) to (n+10), i.e. ODN_5.5 – ODN_5.14. Gel electrophoresis analysis allows a fast and clear evaluation of the overall selectivity of the cross-link reaction depending on the linker used and the position of "T" in the overhang region.

Therefore, we have designed linker molecules in form of building blocks that can be easily attached to DNA in a one-step synthesis. We focused our interest on short and linear linker to obtain greatest selectivity due to stringent localization in space. Importantly, the synthesis of the reactive sequences should occur under mild conditions in aqueous solution, and should not interfere with the functional groups of DNA. In line with this, all linkers were designed as diamines, fused to DNA and modified with N-succinimido-activated bromoacetate as described in **Scheme 5.2**. Examination of cross-linking efficiency was conducted by PAGE analysis.

We first studied cross-link formation by thymidine alkylation using a reactive sequence with only a short ethylenediamine linker (SRS_5.4) (**Fig. 5.7**). Gel electrophoresis results showed that the single band with lower electrophoretical mobility was obtained with the primer that contained thymidine in position (n+1) to the reactive group (ODN_5.5) with efficiency of 27% (lane 2). A band with slightly weaker intensity (yield is equal to 12%) was observed with thymidine in position (n+2) (ODN_5.6) (lane 3). Only traces of product were detected with target nucleotides that were placed further away from the reactive group. Control reactions with the primer only, without addition of reactive sequence, showed no cross-link formation.

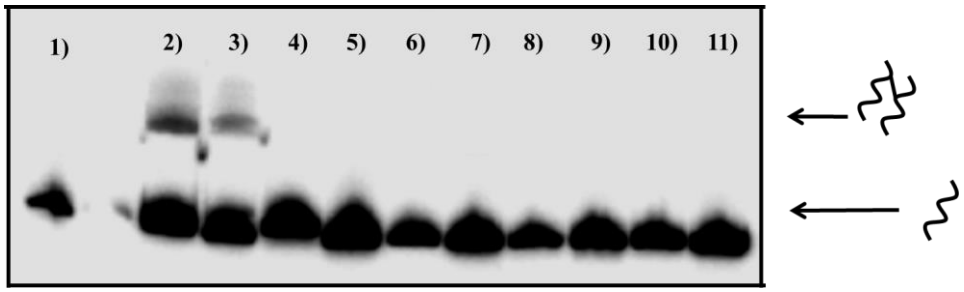


Figure 5.7. Phosphorimage autoradiogram of 18% denaturing PAGE analysis of alkylation of thymidine with a 10-mer donor molecule (SRS_5.4), which contains an ethylenediamine linker and complementary templates with T residues at varying distances from the reactive group (ODN_5.5 - ODN_5.14). Incubation of template ODN_5.5 in absence of the reactive sequence is shown in lane 1.

Incorporation of a cleavable group in the linker of the reactive sequence significantly increases the linker size compared to the ethylenediamine moiety and hence changes of the site-specificity of modification are expected. To mimic the longer linker we used hexylenediamine as linker (SRS_5.5), which allows a more realistic model for cross-link formation.

After evaluation of the reaction with PAGE we observed that the hexylendiamine linker shows the same specificity as the short ethyledediamine linker (**Fig. 5.8**). The DNA templates in

which the T residues were closest to the duplex region (ODN_5.5 and ODN_5.6), (lanes 2 and 3) were covalently cross-linked with efficiencies of 17% and 15%, respectively.

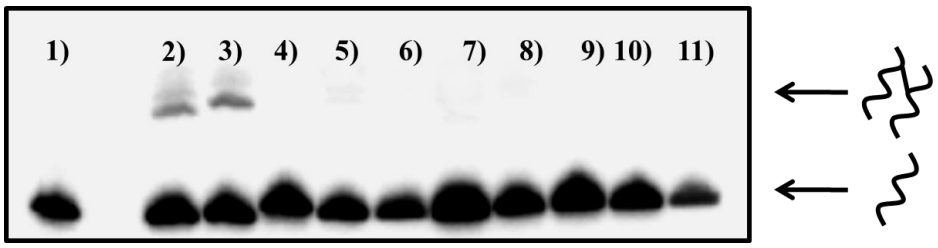


Figure 5.8. Phosphorimage autoradiogram of 18% denaturing PAGE analysis of alkylation of thymidine with 10-mer donor molecule, which contains hexylenediamine linker (SRS_5.5) and complementary templates with T residues at different distance from reactive group (ODN_5.5 - ODN_5.14). In control experiment SRS_5.5 was annealed with mismatched sequence with target guanine at position n+1 (ODN_4.4) (lane 1).

Next we investigated formation of cross-links using reactive sequences with cleavable linker. For this we chose disulfide-containing cystamine because of its stability under a variety of conditions and the mild conditions required for cleavage which are compatible with nucleic acid chemistry. In chapter 4, we demonstrated the application of the disulfide bridge to guanine-derivatized DNAs and optimized the cleavage reaction conditions in order to avoid DNA degradation. Cross-link experiments with the cystamine-containing reactive sequence SRS_5.6 were performed by mixing the bromine-modified reactive sequence and the complementary oligonucleotides (ODN_5.5 - ODN_5.14) in a ratio of 2:1 in $NaHCO_3$ buffer at pH = 9.3 (**Fig. 5.9**). The efficiency of cross-link formation was investigated by PAGE analysis. It indicated that the target template was modified by the disulfide-containing SRS with approximately the same efficiency and specificity as sequences without cleavable linkers (15% for T in (n+1) position and 12% for T in (n+2) position). However, in contrast to previous results also minor cross-link formation was observed for templates with T residues in position (n+4) and (n+5) (3% and 5% respectively) (lanes 5 and 6). These results seem to be consistent with our model proposal based on the linker size study form chapter 3. The selectivity in cross-link formation decreased with increasing size of linker that connected recognition sequence and reactive group.

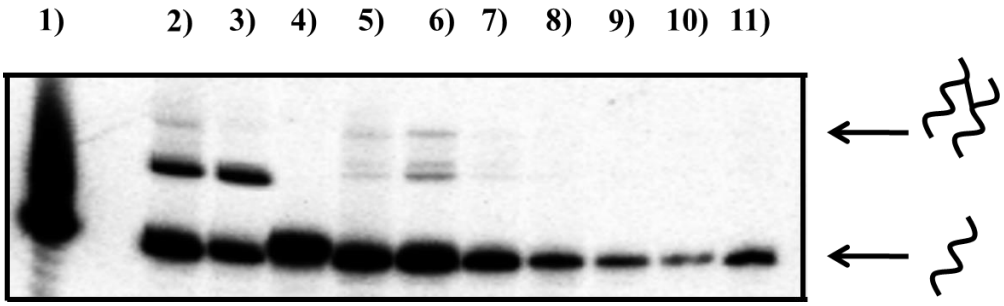


Figure 5.9. Phosphorimage autoradiogram of 18% denaturing PAGE analysis of alkylation of thymidine between 10-mer reactive sequence, which contains cystamine cleavable linker (SRS_5.6) and complementary templates with T residues at different distance from reactive group (ODN_5.5 - ODN_5.14). In control experiment donor molecule was annealed with mismatched sequence with guanine at position n+1 (ODN_4.4) (lane 1).

In the next step we investigated the possibility of thymidine alkylation within the duplex region. In contrast to the N7 position of guanine, the N3 position of thymidine is involved in Watson-Crick pairing with its complementary adenine base. Accordingly, the position might not be accessible for alkylation. To evaluate thymine alkylation within the duplex a longer 17-mer reactive sequence (SRS_5.7) was used with ODN_5.7, i.e. where thymidine is in the (n-5) position, with ODN_5.9 thymine in (n-3) position, as well as ODN_5.12 as a control with thymine in (n+1) position. The 17-mer reactive sequence contains either a diaminohexane (SRS_5.8) or a cystamine linkers (SRS_5.9) (**Fig. 5.10**). In case of diaminohexane linker formation of secondary structure prevents the N3 position of thymidine from alkylation (lane 2 and 3). Cross-link formation was only observed with ODN_5.12 with T in position (n+1) with efficiency of 14%. In contrast, a low amount of cross-linked product is observed for the disulfide-containing reactive sequences (SRS_5.9) with yields of 2% and 4% respectively. The observed results can be explained when viewing the DNA helix as a dynamic structure. At a certain point in time, the chains may separate and the Watson-Crick faces of the bases become accessible.

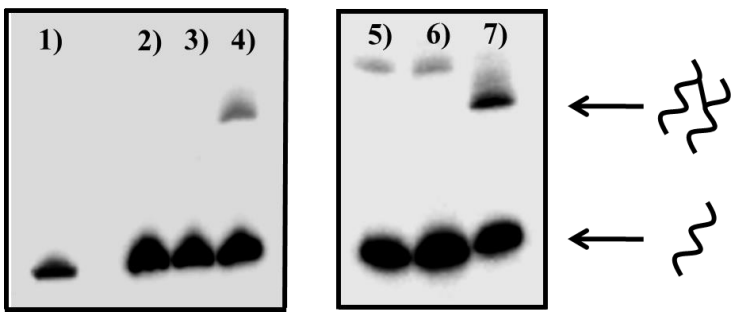


Figure 5.10. Phosphorimage autoradiogram of 18% denaturing PAGE analysis of alkylation of thymidine between 17-mer donor molecule, which contains hexylamine (SRS_5.8) (lanes 2-4) or cystamine (SRS_5.9) (lanes 5-7) cleavable linker, with complementary template with T residues in position (n-5) (ODN_5.7) (lanes 2 and 5), (n-2) (ODN_5.9) (lanes 3 and 6) or (n+1) (ODN_5.12) (lanes 4 and 7). In control experiment reactive molecule (SRS_5.8) was annealed with mismatched sequence with target thymidine at position n+1 (lane 1).

5.7 Dependence of cross-link formation efficiency from pH value

Examination of the alkylation of thymidine monomers demonstrated high yield in alkaline conditions and lower conversion of thymidine with decreasing pH until reaction stops at neutral pH values. Due to the fact that oligonucleotide chains are sensitive to phosphodiester hydrolysis at elevated pH[14], we decided to analyze the reaction at an intermediate pH value. Here we used a 15-mer recognition sequence (SRS_5.10) to preserve DNA helical structure under alkaline conditions. Reactive molecule also contains cystamine linker and the usual 2-bromoacetate reactive group (SRS_5.11). Cross-link formation was tested with complementary template ODN_5.15 at different pH values (**Fig. 5.11**). We observed, that yields only change slightly (around 5%) when the pH was lowered from 9.3 to 8.6. Assuming, that oligonucleotides are more stable at less alkaline buffers[21],[22],[23], for the future applications we accepted a slightly lower yield for the sake of milder reaction conditions.

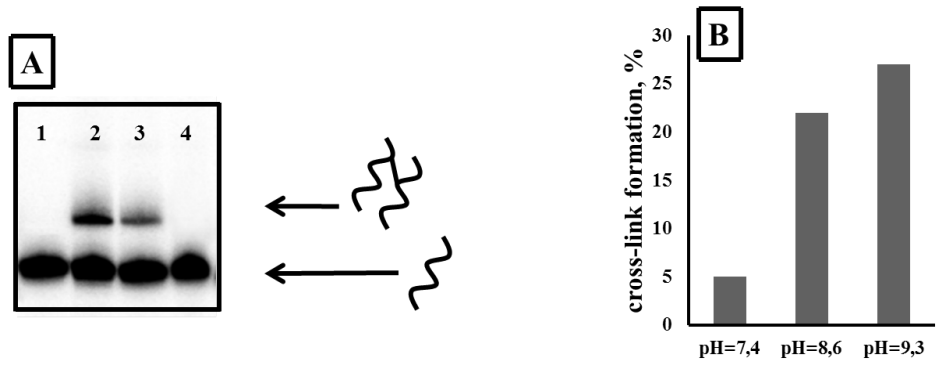
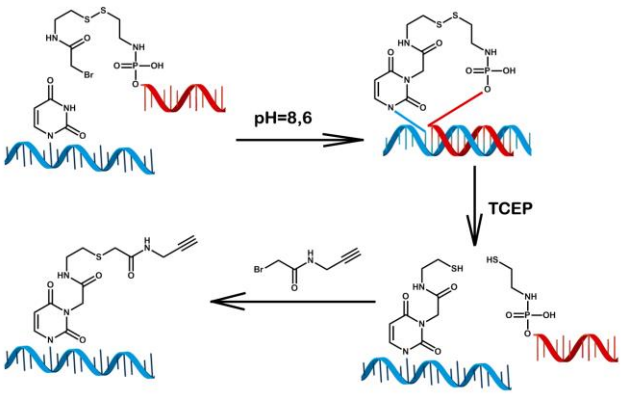


Figure 5.11. **A**: Phosphorimage autoradiogram of 12% denaturing PAGE analysis of alkylation of thymidine between 15-mer donor molecule, which contains cystamine cleavable linker(SRS_5.11), with complementary template with T residue at position n+1 (ODN_5.15) at pH = 9.3 (lane 2), pH = 8.6 (lane 3) and pH = 7.4 (lane 4). In the control experiment reactive sequence was annealed with mismatched sequence with target thymidine at position n+1 (lane 1). **B**: Comparative graph of intensity of cross-linked products between reactive sequence and complementary template at different pH values.

5.8 Application of N3-alkylation of thymidine for attachment of a functional group

Our main goal was to use the alkylation of thymidine with cleavable linker for the introduction of a functional group. Hence we reacted SRS_5.11 with complementary sequence ODN_5.15 that contains T residues in (n+0), (n+1) and (n+4) positions at pH = 8.6. Concerning our previous results, we assumed that only thymines in (n+0) and (n+1) positions to the reactive group will be modified. At the same time, distance of four nucleotides from reactive group to T in (n+4) position almost inactivates reaction of alkylation and we can neglect this side reaction. The mild chemical cleavage conditions are expected to yield thiol modified target DNA sequences which can be further used for modifications or labelling.

However, trapping of the free thiol-group after reductive cleavage of disulfide containing linker with TCEP proved to be more difficult than expected[24]. First we tried the well-known Michael coupling between newly formed thiol group and 2-bromo-N-(propargyl)acetamide which gave no product. Next we tried the alkylation of the thiol-group with a bromoacetate-derivative. For this reason we synthesised a water soluble amide derivative of bromoacetate with a triple bond in the side chain (see experimental part). This molecule was then used for thiol-group alkylation in template DNA resulting in formation of a stable and inert alkyne-group in the DNA template (**Scheme 5.3**). After RP-HPLC purification MALDI-TOF MS demonstrated the presence of initial unmodified DNA template and modified DNA (**Fig. 5.12**). According to HPLC analysis, the alkyne-modified DNA template was obtained in 7% yield.



Scheme 5.3. Site-specific incorporation of thiol-group in template DNA by alkylation of thymidine N3 position followed by cleavage of the disulphide linker and catching the newly formed thiol-group with 2-bromo-N-(propargyl)acetamide.

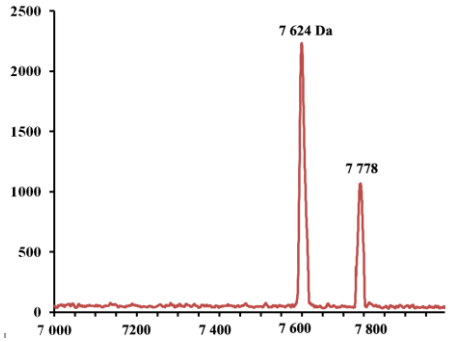
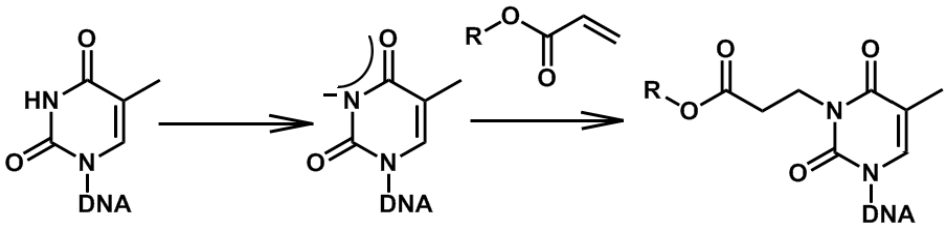


Figure 5.12. MALDI-TOF spectrum of the alkyne-containing modified DNA template. Molecular mass of unmodified DNA template is $M_{calc.}$ = 7627.3 Da, modified template $M_{calc.}$ = 7774.9 Da.

5.9 Second type of specific-reaction: nucleophilic addition of thymidine to an activated multiple bond

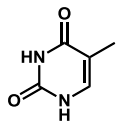
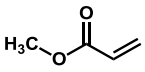
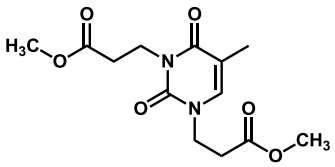
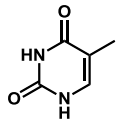
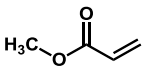
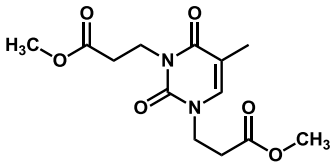
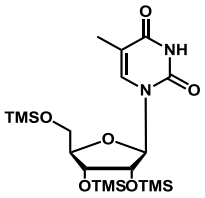
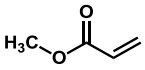
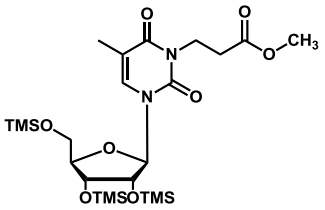
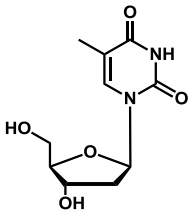
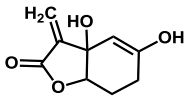
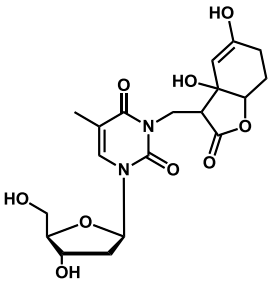
Besides the capability of the thymine N3 position to be alkylated, deprotonated N3 can also undergo a nucleophilic addition to an activated multiple bond at alkaline conditions (**Scheme 5.4** and **Table 5.3**). To expand the tools for point-specific modification we tested the nucleophilic addition of four oligonucleotides to the acrylamide double bond under alkaline conditions. Incorporation of the acrylamide side chain resulted in the formation of a new peak during reverse phase HPLC which was detected only in the reaction with thymidine residues (see Experimental part). The pH-yield dependence demonstrated conversion of thymidine only under basic conditions, and no reaction occurred at neutral pH (**Fig. 5.13**).



Scheme 5.4. Alkylation of thymidine with acrylic acid derivatives under alkaline conditions.

Table 5.4. Examples of alkylation thymine at N3 position.

Entry	Thymine derivative	Conditions	Product	Ref

1		DBU, DMF, 1 day, RT 		19
2		DBU, DMF, 2 hours, 80°C 		13
3		DBU, DMF, 1 day, RT 		17
4		NaOH, H_2O-THF, 16 hours, RT, pH = 8.7 		5

The reaction was analyzed by LC-MS and electrospray ionization mass spectroscopy (**Fig. 5.13**). UV-trace of the reaction mixture under alkaline conditions of acrylamide with thymidine showed the presence of the new peak of molecular ion of 337.1 Da in the positive and 312.9 Da in the negative mode during mass spectrometric identification.

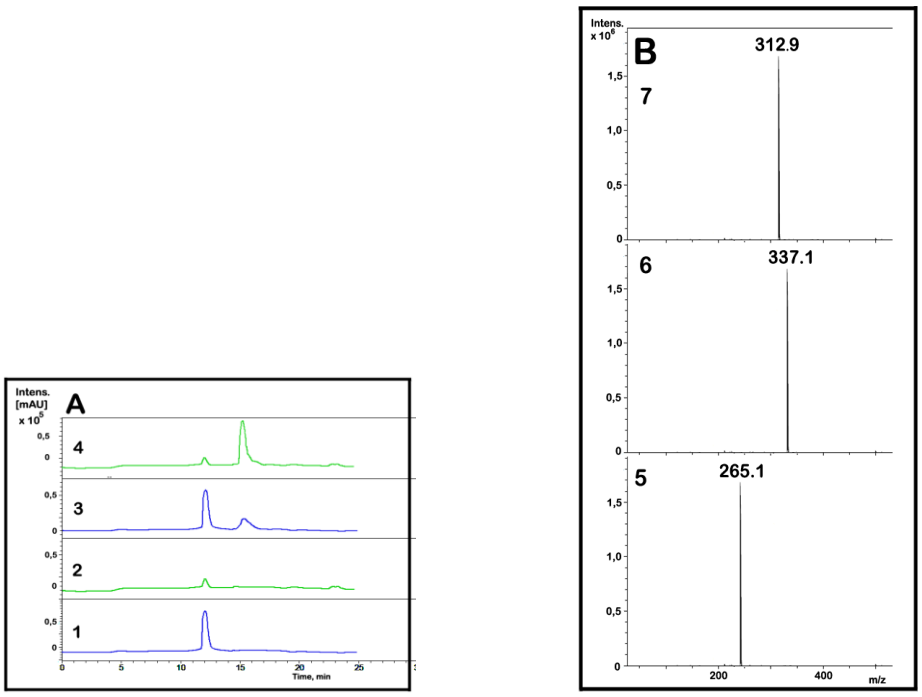


Figure 5.13. **A**: Reverse-phase HPLC chromatogram with different detectors (λ = 260nm – blue line, ESI negative mode – green line) of **1,2**: reaction of thymidine with bromoacetate at pH = 5.6; **3,4**: reaction of thymidine with bromoacetate at pH = 9.3. **B**: Deconvoluted ESI-MS spectra of **5**: Thymidine eluted at 12.5 min ($[M+H+Na]^+_{calc.}$ mass 265.2 Da), **6**: modified thymidine eluted at 15.8 min in positive mode ($[M+H+Na]^+$calc. 337.16 Da), **7**: modified thymidine eluted at 15.8 min in negative mode ($[M-H]^-$calc. 312.9 Da).

10-mer (SRS_5.12) and 17-mer (SRS_5.13) reactive sequences with cystamine linker and acrylamide moiety as the reactive group were synthesized similarly to the previously described protocol. Both reactive sequences were annealed with complementary sequences in which the target thymine was placed in the single strand region at different distances from the reactive group (ODN_5.5 – ODN_5.14). In case of the 17-mer SRS_5.13 the target thymine was placed in the single strand region (ODN_5.12) or in the double strand region (ODN_5.7 and ODN_5.9). Incubation of DNA templates with acrylamide-modified reactive sequences at pH = 8.6 was expected to result in the formation of a covalent cross-link *via* nucleophilic addition and hence resulting in formation of a new band with lower electrophoretic mobility (**Fig. 5.14**). Surprisingly, we did not observe any cross-link formation with thymidine which was placed in the single strand region. However, clear cross-link formation was observed with ODN_5.7 and ODN_5.9 (lanes 12, 13 and 16), where thymidine was incorporated in the duplex region.

Such a result can be interpreted by the nucleotide in duplex region being strongly fixed in space by the hydrogen bond with complementary base and π-stacking with neighboring nucleotides. The molecular orbital of the N3 nitrogen atom is also strongly localized. The amide

of the acrylic acid is hydrophobic and may be incorporated in the middle of DNA strand. This strong localization of two molecular orbitals in space close to one another may facilitate nucleophilic addition to target thymidine and result in the site-specific formation of the cross-link band. In the case of thymidine in the single strand it was impossible to create such strong localization of the 3D structure and any attempts of cross-link formation failed.

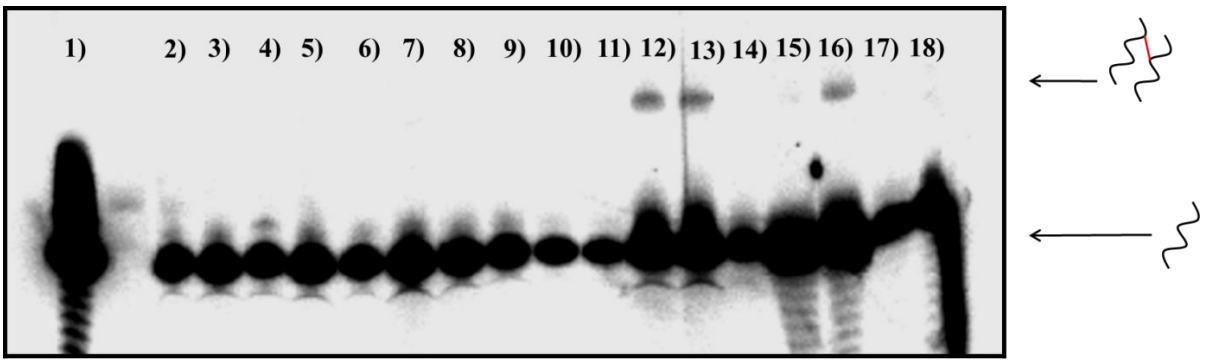


Figure 5.14. Phosphorimage autoradiogram of 18% denaturing PAGE analysis of alkylation of thymidine between 10-mer reactive sequence SRS_5.12 (lanes 2-11) or 17-mer reactive sequence SRS_5.13 (lanes 12-18), which contains cystamine cleavable linker, with complementary templates with T residues in the single strand (ODN_5.5 – ODN_5.14) (lanes 2-11) or double strand regions ODN_5.7 (lanes 12 and 15), ODN_5.9 (lanes 13 and 16) and ODN_5.12 (lanes 14 and 17). In the control experiment donor molecule was annealed with a mismatched sequence with target thymidine at position n+1 (lanes 1 and 18).

5.10 “Locked” modular system

In chapter 3 we described a construct, in which two duplex regions flanked a single stranded loop region with the reactive group positioned in the middle. Attempting to use a similar locked construct for the alkylation of thymine we used a reactive sequence with an ethylenediamine linker between the reactive bromoacetate-group and the central cytosine base of the reactive sequence (SRS_5.15). Efficiency of cross-link formation at different pH values demonstrated the expected trend with a yield of 40% at alkaline pH = 9.3. Such as RNA oligonucleotides preserve it stability at pH = 8.6, this approach can be also used for the site-specific alkylation of uracil in RNA. In all cases, “locked” approach demonstrated higher yield than the original construct and was enough for further analyses.

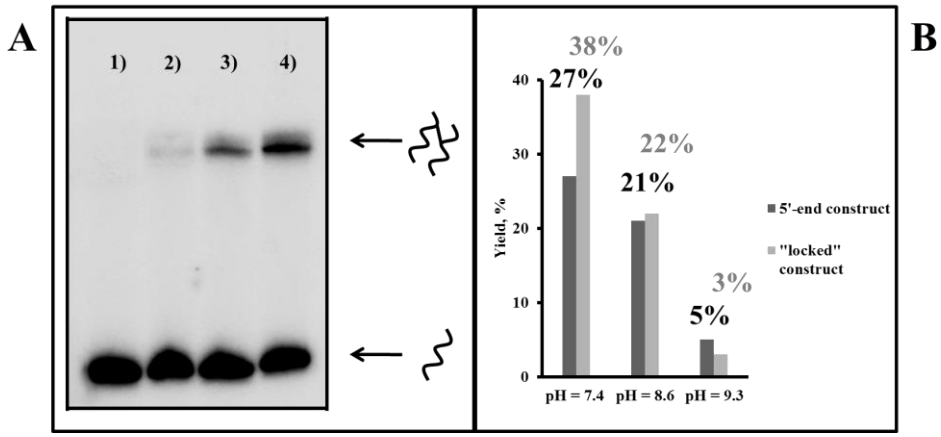


Figure 5.15. **A**: Phosphorimage autoradiogram of 12% denaturing PAGE analysis of alkylation of thymidine with the "locked" construct, which contains an ethylenediamine linker (SRS_5.15), with complementary template with T residue at (n+1) position (ODN_5.16) at pH = 7.4 (lane 2), pH = 8.6 (lane 3) and pH = 9.3 (lane 4). Incubation template in the absence of the reactive sequence is shown in lane 1. **B**: Comparative graph of yields of cross-link products between 5'-end construct with 15-mer reactive sequence SRS_5.11 and ODN_5.15 (**Fig. 5.11**) as well as with the "locked" construct using reactive sequence SRS_ 5.15 and ODN_5.16 at different pH values.

5.11 Determination of cross-linking site with enzymatic digestion

In order to confirm the site-specific formation of cross-linked products with the target thymine base, enzymatic degradation studies using exonuclease III (EXO III) were performed with the cross-linked species, i.e. reaction of SRS_5.15 with ODN_5.16, to allow the localization of the exact position of the cross-linking bond. **Figure 5.16** shows the RP–HPLC spectrum obtained after enzymatic digestion of the cross-linked duplex. Identification of peaks were performed with ESI-MS (**Fig. 5.17**). The signal at m/z 553.1 Da ($M_{calc.}$ = 552.22 Da) corresponds to the protonated $[M+H]^+$ cross-linked product C-T. Cross-link formation seems to influence digestion of the 5'-phosphodiester bond to the cross-link in the complementary sequence, as also a C/TA fragment is observed (M_{exper} = 863.9 Da, M_{calc} = 865.9 Da).

Although the described experiments yield information about the exact location of the cross-link and the selectivity of our construct, no direct structural information as to the exact chemical structure of the cross-linked species can be deducted.

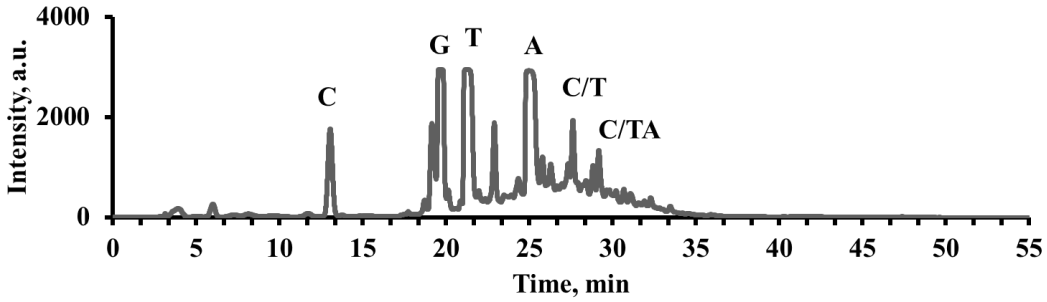


Figure 5.16. Reverse-phase HPLC chromatogram after EXO III digestion of the cross-linked "locked" construct (SRS_5.15) with the complementary template (ODN_5.16). Enzymatic degradation was performed using 200 units of EXO III for desalted cross-linked duplex (0.6 nmol). Peak at 13.5 min responsible for C, 19.2 min for G, 21.8 min for T, 25.1 min for A, 27.0 min – C/T cross-link, 28.3 min for C/TA. Digestions were carried out at 37°C for 50 min. Collected fractions were analyzed by ESI MS (**Fig. 5.17**).

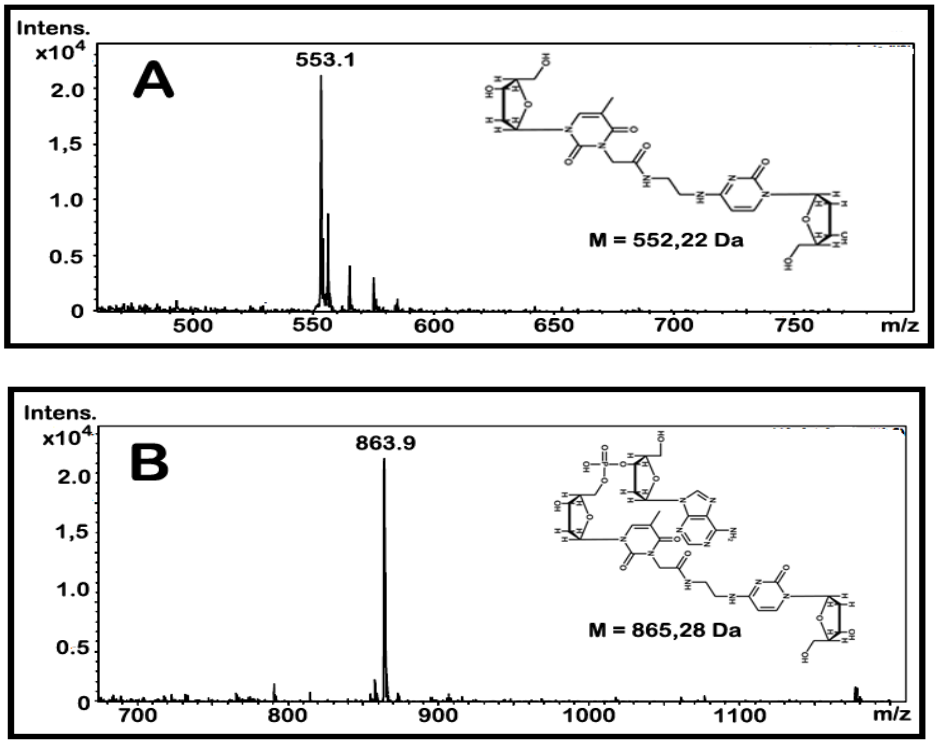


Figure 5.17. Deconvoluted ESI-MS spectrum after enzymatic digestion of the cross-linked compound. **A:** ESI-MS in positive mode of fraction that eluted at 27.0 min (C/T), M_{calc} = 552.22 Da. **B:** ESI-MS in negative mode of fraction that eluted at 28.3 min (C/TA), M_{calc} = 865.28 Da.

5.12 Conclusion

The results displayed in this chapter clearly demonstrate that the locked construct is certainly one of the best systems for the mild, fast, selective and site-specific cross-linking of long oligonucleotide sequences. The alkylation of thymidine exhibits high chemoselectivity which gives rise to the possibility of introduction of different functional groups. Reactive sequences can be conveniently synthesized from a DNA oligonucleotide containing a 5'-phosphate group and bromoacetic acid. High selectivity for T in (n+1) and (n+2) positions is observed with ethylenediamine, hexylenediamine or linkers with similar size. Accordingly, site-specific incorporation of a thiol group can be achieved from the cross-link reaction with a disulfide bridge containing linker after reductive cleavage. Finally, the newly formed thiol group was labeled with an alkyne-containing molecule. The stable and relatively inert alkyne group can be used for further labeling with e.g. azide-containing dyes using "click" chemistry.

5.13 Reference

(1) Yildirim, İ.; Akcamur, Y. *Turk. J. Chem.* **1996**, *20*, 27.
(2) Verdolino, V.; Cammi, R.; Munk, B. H.; Schlegel, H. B. *The Journal of Physical Chemistry B* **2008**, *112*, 16860.
(3) Roy, B.; Dutta, S.; Chowdhary, A.; Basak, A.; Dasgupta, S. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 5411.
(4) Hadj-Bouazza, A.; Teste, K.; Colombeau, L.; Chaleix, V.; Zerrouki, R.; Kraemer, M.; Catherine, O. S. *Nucleos. Nucleot. Nucl.* **2008**, *27*, 439.
(5) Sakakura, A.; Takayanagi, Y.; Shimogawa, H.; Kigoshi, H. *Tetrahedron* **2004**, *60*, 7067.
(6) Gu, J.; Wang, J.; Leszczynski, J. *Nucleic Acids Res.* **2010**, *38*, 5280.
(7) Ogorodnikova, N. A. *Journal of Molecular Structure: THEOCHEM* **1993**, *279*, 71.
(8) Shabarova, Z. A.; Bogdanov, A. A.; 1 edition (July 13, 1994) ed.; Wiley-VCH: 1994, p 589.
(9) Gasteiger, J.; Walter, R. D. *Pharm.* **1986**, *41*, 661.
(10) Wood, E. J. *Biochemical Education* **1987**, *15*, 97.
(11) Lide, D. R.; 80-th ed.; Press, C., Ed. 2004, p 2712.
(12) Grimm, G. N.; Boutorine, A. S.; Helene, C. *Nucleos. Nucleot. Nucl.* **2000**, *19*, 1943.
(13) Boncel, S.; Mączka, M.; Walczak, K. Z. *Tetrahedron* **2010**, *66*, 8450.
(14) Cadet, J.; Delatour, T.; Douki, T.; Gasparutto, D.; Pouget, J.-P.; Ravanat, J.-L.; Sauvaigo, S. *Mutat. Res-Fund. Mol. M.* **1999**, *424*, 9.
(15) Chandra, M.; Keller, S.; Luo, Y.; Marx, A. *Tetrahedron* **2007**, *63*, 8576.
(16) Wang, J.; Wang, Y. *Nucleic Acids Res.* **2009**, *37*, 336.
(17) Bonache, M.-C.; Chamorro, C.; San-Félix, A. *Antiviral Chem. Chemother.* **2003**, *14*, 249.
(18) Mabey, W.; Mill, T. *J. Phys. Chem. Ref. Data* **1978**, *7*, 383.
(19) Boncel, S.; Walczak, K. *Lett. Org. Chem.* **2006**, *3*, 534.
(20) Taylor, M. J.; Dervan, P. B. *Bioconjugate Chem.* **1997**, *8*, 354.
(21) Calladine, C. R.; Drew, H. R.; Luisi, B. F.; Travers, A. A. *Understanding DNA: the molecules and how is it work*; third addition ed.; Elsevier Academic Press, 2004.
(22) Yakovchuk, P.; Protozanova, E.; Frank-Kamenetskii, M. D. *Nucleic Acids Res.* **2006**, *34*, 564.
(23) Nakano, S.-i.; Fujimoto, M.; Hara, H.; Sugimoto, N. *Nucleic Acids Res.* **1999**, *27*, 2957.
(24) Dempsey, G. T.; Bates, M.; Kowtoniuk, W. E.; Liu, D. R.; Tsien, R. Y.; Zhuang, X. *J. Am. Chem. Soc.* **2009**, *131*, 18192.

CHAPTER 6

Site-specific incorporation of a pair of Förster resonance energy transfer dyes and their properties

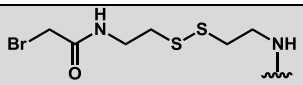
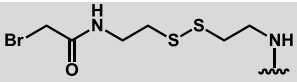
6.1 Introduction

Affinity probes equipped with additional reactive groups for covalent linking, provide a straightforward way for capturing and labeling[1],[2],[3]. The basic probe architecture, which combines multiple functions in one entity, remained largely unchanged since the first studies. It lacks modularity and thus presents several opportunities for further research. First, the reactive group should be distal from the recognition sequence to prevent self-modification but be close enough to preserve site-specificity, i.e. to modify only the target position. Second, for each experiment, preparation of multiple probes serving different purposes is often required, e.g. with a fluorophore for detection as well as an affinity tag for recognition. Finally, the stability of the fluorophore probes during the reaction can drastically change the imaging picture. Previously reported affinity probes were initially designed for one modification at one specific position in only one type of molecule[4],[5].

Here we report one possible application for our modular system. In the previous chapters three independent units, i.e. recognition DNA sequence, the cleavable linker and the reactive group, have been used to assemble the reactive sequence with controlled reactivity for site-specific modification of the target oligonucleotide primer. DNA-templated chemistry was used to establish a simple, modular, and multiplexed system for the labeling of native oligonucleotides with small molecules. In this chapter, we attempt the conjugation of a fluorophore mediated by the modular system as a bioconjugate tool. Site-specific incorporation of two types of cyanine dyes was proved by Förster resonance energy transfer (FRET).

Sequences used:

*Complementary regions are underlined

Name	Reactive group	Sequence
SRS_6.1		5'-AT GCG GCA GGC TGC A-3'
ODN_6.1[a]		5'-TGC AGC CTG CC**G** CAT GAC TCG TGG T -3'
ODN_6.2[b]		5'-(**biotin**)-ACC ACG AGT CAT GCG GCA GGC T(**Cy3**)GC A -3'
SRS_6.2		5'-GTT TAT TCC GCC TAT CAG CC -3'
ODN_6.3[a]		5'- GAA GGC AAG CGG CTG ATA GGC GGA ATA AAC **G**AC TGA AGA G -3'
ODN_6.4[b]		5'-**Cy5**-TTC TTG TTA TTG CGT TCC ATC CTA TCA GCC GCT AGT GAG C –(**biotin**)-3'

[a]: Target templates.

[b]: Complementary templates for FRET experiment.

6.2 Overview: Förster resonance energy transfer

The development of molecular biology and gene therapy requires specific and sensitive detection of nucleic acid sequences identified as disease markers[6],[7]. Fluorescence technology is one of the perspective tools for the visualization, interaction and modification of targets in living cells. Until now three strategies for biomolecular imaging were used.

1) Fluorescent dyes with a reactive group are possible candidates for biomolecular imaging however they usually bind to multiple targets[8],[9]. In contrast to our modular system, they non-specifically modify target sequences as the reactive groups used, e.g. NHS, maleimide and hydrazide, is not selective (**Fig. 6.1**). Organic fluorophores contain long conjugated π-electron systems for absorption in the visible or near-visible spectrum and due to it usually have a low solubility under physiological conditions. Incorporation of charged substituents such as sulfonic or carboxylic acids partially increases the solubility, but also decreases the photostability of the fluorophore[10]. Inspired of photobleaching and background fluorescence issues, this approach was used in a number of applications[11].

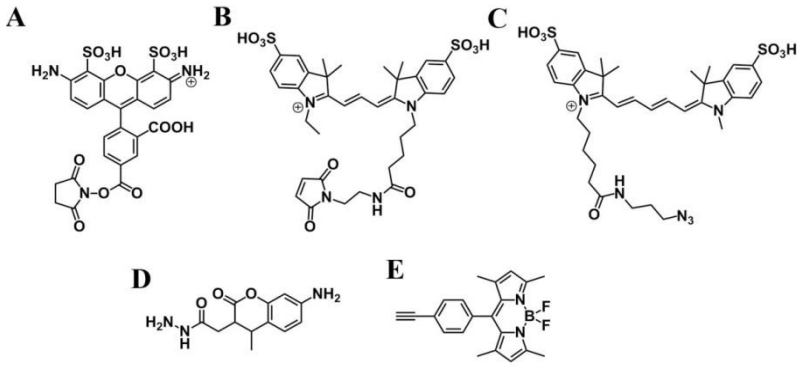


Figure 6.1. Fluorescent dyes used in FRET application: **A**: AlexaFluor 488, succinimidyl ester (Life Technologies Europe B.V.); **B**: maleimido-sulfo-Cy3 (GE Healthcare); **C**: azido-sulfo-Cy5 (Lumiprobe); **D**: AMCA-hydrazide (Pierce Biotechnology, Inc.); **E**: alkyn-BODIPY (Life Technologies Europe B.V.).

2) Another approach for the visualization of biomolecules is to inject fluorescently labeled artificial sequences that selectively bind to the target sequence and produce a signal. Tyagi and colleagues[12] developed such anti-sense sequences, which they called "molecular beacon" (MB). These probes includes DNA sequence for the recognition of specific sequence marker in target DNA and fluorophores[13],[14] or quenchers[15] at both ends of recognition sequence.

3) The fully biological approach includes incorporation of a vector that encodes for the green fluorescent protein (GFP) into the cell and replication of this protein inside the cell. Shav-Tal et al. attached GFP-sequence to an mRNA molecule that codes for the b-actine peptide and after translation b-actine with attached GFP at one end was obtained[16].

Different fluorescent techniques have been described and include the direct excitation of the fluorophore without emission control[3],[17]. In contrast, Förster resonance energy transfer requires a pair of fluorophores, donor and acceptor (**Figure 6.2**). After excitation, the excite-state energy level of the donor fluorophore can be non-radiatively transfered to an acceptor fluorophore. This transfer results in quenching of the donor fluorophore emission and enhancement of acceptor fluorescence intensity. Energy transfer from the donor to the acceptor can occur efficiently when both fluorophores are placed at a short distance from each other and decreases with increasing distance between them (inverse of sixth power of the distance). As a result, binary FRET probes can be activated only when both fluorophores are introduced at specific positions in the target template and annihilate any non-specific modifications[18]. FRET has been widely used to detect association of proteins[19], protein conformational changes[20], and hybridization between donor-labeled oligodeoxynucleotides and acceptor-labeled complementary oligodeoxynucleotides[21].

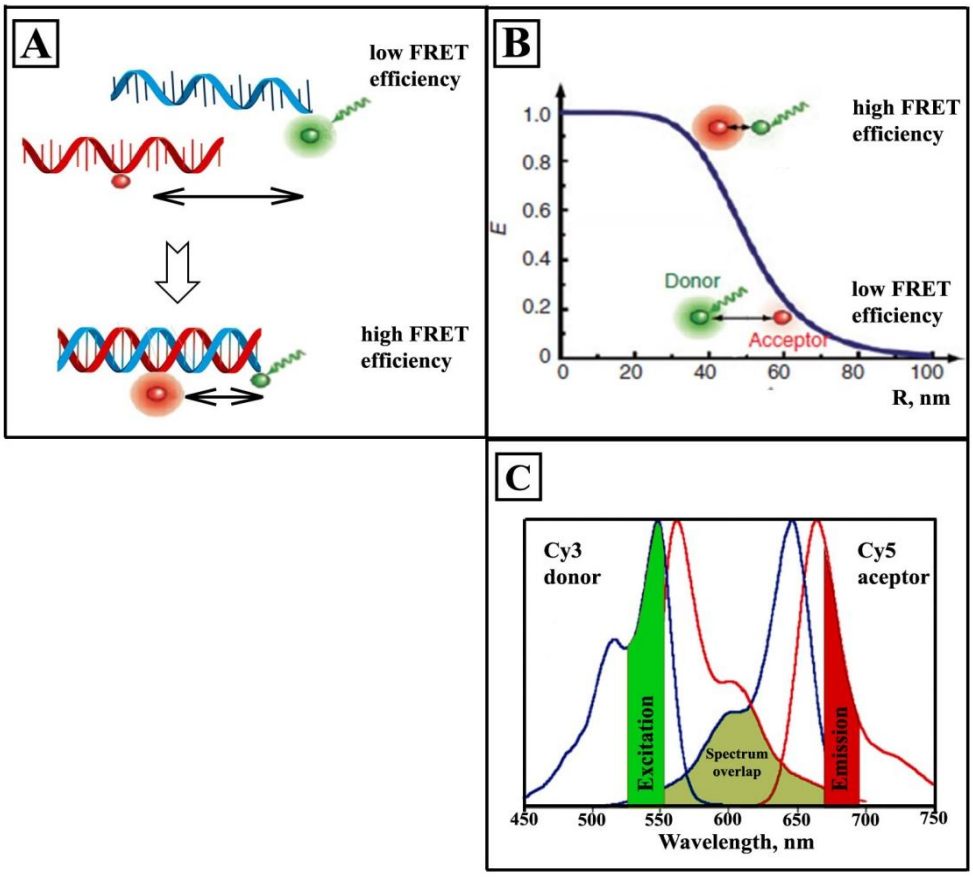


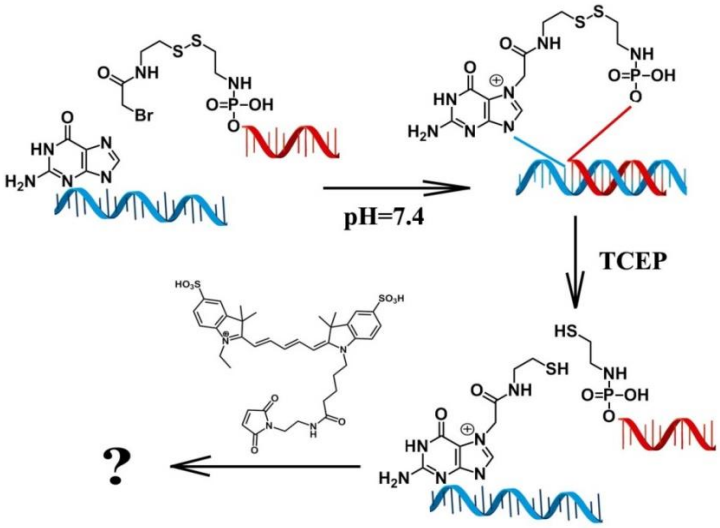
Figure 6.2. Principles of Förster resonance energy transfer (FRET). **A**: Two DNA primers containing fluorophores (green bulb – donor, red bulb – acceptor) that are separated by the distance R can undergo energy transfer upon excitation (green arrow) of donor. Emission of donor fluorophore is shown with a green sphere around the bulb, emission of acceptor (caused by energy transfer) with a red sphere. Localization of two fluorophores in space is done due to annealing complementary oligonucleotide strands. **B**: Distance dependency of energy transfer efficiency **C**: Excitation (green line) and emission (red line) for donor (Cy3) and acceptor (Cy5) fluorophores. The spectra are normalized for comparison. Excitation wavelength of donor for FRET application is shown in green. Emission of acceptor due to energy transfer is shown in red. The spectral overlap of donor emission and acceptor excitation is shown as light green area under the curve. Picture was adapted from (http://www.biotek.com/resources/articles/fluorescence-resonance-energy-transfer.html) published by BioTek Instruments, Inc., June 20, 2005.

6.3 Site-specific labeling with acceptor (sulfo-Cy5) fluorophore

To determine the efficiency of site-specific dye incorporation we designed an oligonucleotide-based DNA substrate with a pair of fluorophores attached to the DNA double helix. We purchased a single strand 25-bp oligonucleotide DNA substrate with a donor cyanine dye and biotin (ODN_6.2), which gives the opportunity for immobilization of the strand on the surface to facilitate energy transfer. The acceptor dye was introduced to the unmodified complementary single strand DNA, the target strand, using the post-synthetic approach described in Chapter 4 (**Scheme 6.1**). The reaction is based on the site-specific alkylation of the N7 position of the target guanine with a 2-bromoacetate-containing DNA reactive sequence complementary to the specific region.

The synthesis of reactive sequence containing the disulfite cleavable moiety was accomplished as described in the previous chapter[22]. Reaction of activated oligonucleotide with cystamine in the presence of 2,2'-dithiopyridine and triphenylphosphine gave phosphoramidite DNA with the cleavable linker which was converted to the final reactive sequence (SRS_6.1) by acetylation with the N-hydroxysuccinimine ester of 2-bromoacetate. The yield of reaction is 78% relative to the starting compounds. Because of concern about bromoacetate group stability toward base induced hydrolysis, it was immediately diluted in the hybridization mixture and left for hybridization for 24 h.

To investigate the feasibility of the method, we first used the reactive sequence for cross-link formation with a short radiolabeled DNA molecule of 25 nucleotides (ODN_6.1). The target guanine was placed in 10 nucleotide (30 nm) distance from the Cy3 modified end of the complementary sequence. With this distance the efficiency of energy transfer between the Cy3-Cy5 pair of fluorophores should reach around 70%[23]. Cross-linking experiments were conducted after annealing of both strands. Stability of the duplex between SRS_6.1 and the target sequence ODN_6.1 plays a critical role in cross-link formation. As predicted from donor – template duplex stability ($T_{melting}$ = 53°C), the yield of the cross-linked product should reach approximately 15%.



Scheme 6.1. Site-specific modification of target DNA strand (blue curve) with modular system (red curve). The newly introduced thiol group in the target oligonucleotide is labelled by maleimide-sulfo-Cy5 fluorophore.

The reaction was performed in different buffers: MOPS (pH = 7.4), sodium borate (pH = 8.6) and sodium carbonate (pH = 9.3) (**Fig. 6.3.**). As expected the highest yield of cross-link formation was obtained at the higher pH value and decreased in the order carbonate> borate = MOPS. Nevertheless, N7-alkylated guanine demonstrated higher stability at neutral pH, so further reactions were performed in MOPS buffer at pH = 7.4.

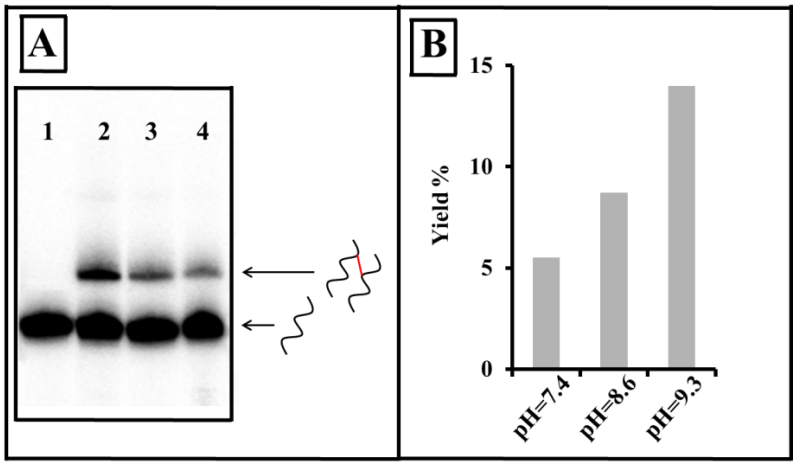


Figure 6.3. Phosphorimage autoradiogram of 12% denaturing PAGE analysis of cross-link formation between SRS_6.1 containing cleavable cystamine linker and 2-bromoacetate as reactive group and template ODN_6.1. **A**: Reaction was done at pH = 9.3 (lane 2), pH = 8.6 (lane 3) and pH = 7.4 (lane 4). In the control experiment template primer was incubated in the absence of reactive sequence (lane 1). **B**: Comparative graph of cross-linked product yield between SRS_6.1 and template ODN_6.1 at different pH values.

After incubation for 24 hours, the disulfide bridge between the reactive and the target DNA sequences was reduced with TCEP leading to the formation of a thiol group that serves as an anchoring point for a number of functional groups. To demonstrate the labeling principle, we selected a maleimide-functionalized fluorophore that selectively reacts with thiols. Incorporation of the Cy5-dye was conducted immediately after reduction of the disulfide bridge for 6 hours. It is important to mention that the cyanine dye could react with both, the thiol group in the target DNA as well as the thiol group in the recognition sequence. As a control, maleimide-sulfo-Cy5 was reacted with the reduced reactive sequence in the absence of the target strand and used as a standart in PAGE and RP-HPLC analyses of the original reaction.

Fluorescent gel electrophoresis analysis gave a quick and clear idea of the overall selectivity of template labeling in terms of separation of the reactive and template strands and the cross-linked product (**Fig. 6.4**). In the negative control experiment we incubated the cyanine dye with the unmodified template and the duplex strand. SDS-PAGE analysis did not present significant changes suggesting that maleimide-sulfo-Cy5 is inert towards reaction with unmodified single or double stranded oligonucleotides. We then performed cleavage of the disulfide bridge and formation of the conjugate between the reactive sequence and sulfo-Cy5 fluorophore. The formation of a new characteristic band was observed and assigned to the Cy5-labeled reactive sequence. Following incubation of the reactive and target sequences at room temperature, we observed in addition to the Cy5-labeled reactive sequence the presence of a band with slower migration in a 18% denaturing PAGE gel, indicative of successful template labeling.

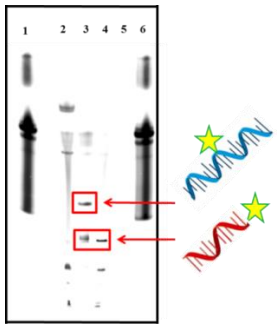


Figure 6.4. Fluorescent scan of denaturing 18% PAGE, showing analysis of DNA labeling. The fluorophore-coupled target DNA and the reactive sequence are indicated on the right-hand side of the image. **Lane 1**: initial non-modified recognition sequence was annealed to target template primer followed by treatment with maleimide-sulfo-Cy5. **Lane 2**: incubation of single stranded target DNA template with maleimide-sulfo-Cy5, **Lane 3**: site-specific labeling of target template with reactive sequence and maleimido-sulfo-Cy5, **Lane 4**: labeling single stranded reactive sequence in the absence of target template. **Lane 5**: loading buffer with template oligonucleotide. **Lane 6**: negative control with only the maleimide-fluorophore present.

As described above, reduction of the cleavable linker gives two modified oligonucleotides that are different in size. Indeed, RP–HPLC spectral analysis at λ = 260 nm shows the presence of the initial unmodified product, as well as two additional peaks with lower retention time (35.6 min and 39.9 min) (**Fig. 6.5. A and C**). Spectral scanning of these peaks demonstrated second absorption maxima at 650 nm, which can be correlated with the dye moiety (**Fig. 6.5. B and D**). The observed retention time for peak at t = 35.6 min was consistent with the experimental retention time of labeled recognition sequence that was independently obtained by conjugation of the donor molecule with Cy5 fluorophore in the absence of the target template.

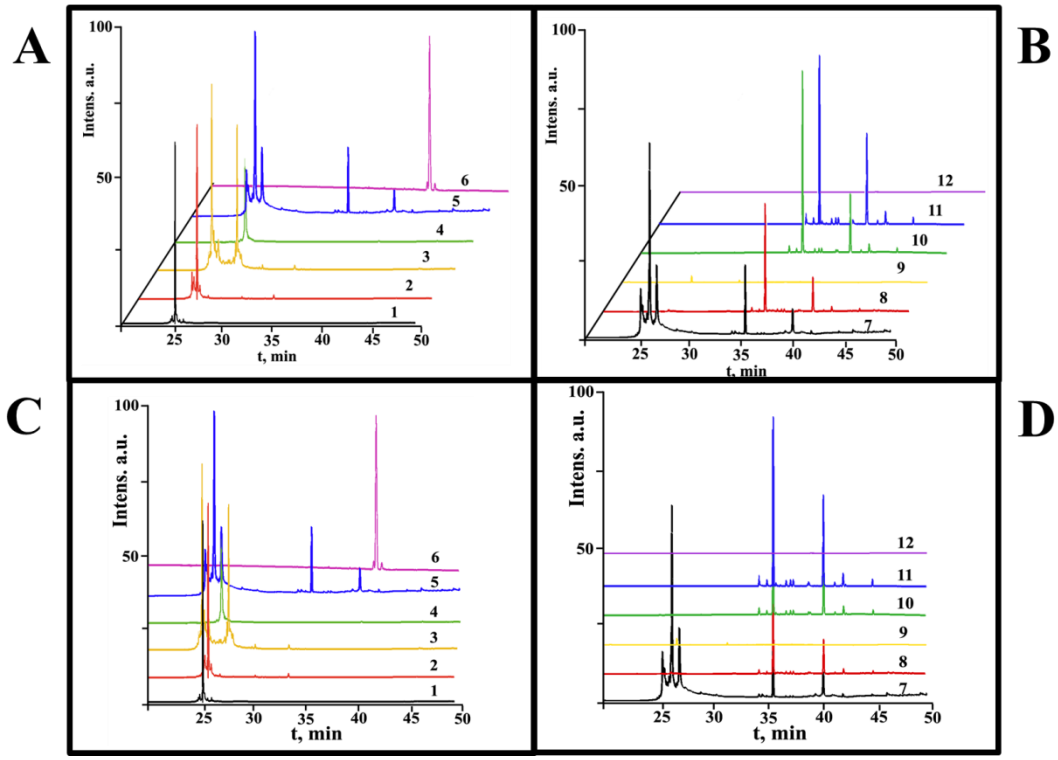


Figure 6.5. Reverse-phase monitoring of labeling of template. **A**: RP-HPLC chromatograms of reaction steps (λ = 260 nm): **1**: initial recognition sequence, **2**: Recognition sequence with cystamine moiety, **3**: Reactive sequence SRS_6.1, **4**: initial target oligonucleotide ODN_6.1, **5**: Sulfo-Cy-5-labeled reaction mixture, **6**: only maleimido-sulfo-Cy5 dye. **B**: Spectral scanning of reaction mixture at **7**: λ = 260 nm, **8**: λ = 320 nm, **9**: λ = 400 nm, **10**: λ = 560 nm, **11**: λ = 650 nm, **12**: λ = 720 nm. **C:** same as **A**; **D:** same as **B**.

The new slower running species (at 35.6 min and 39.9 min) were purified by HPLC and their masses were determined by MALDI-TOF MS analysis. The observed molecular weight (M_{found} = 6343.6 Da) of peak that eluted at 35.6 min was found to be consistent with the calculated mass of sulfo-Cy5 labeled reactive sequence ($M_{calc.}$ = 6346.3 Da). The molecular

weight of second peak at 39.9 min (M_{found} = 8543.6 Da) was corresponded to the sulfo-Cy5 labeled template ($M_{calc.}$ = 8548.4 Da) (**Fig. 6.6. B**).

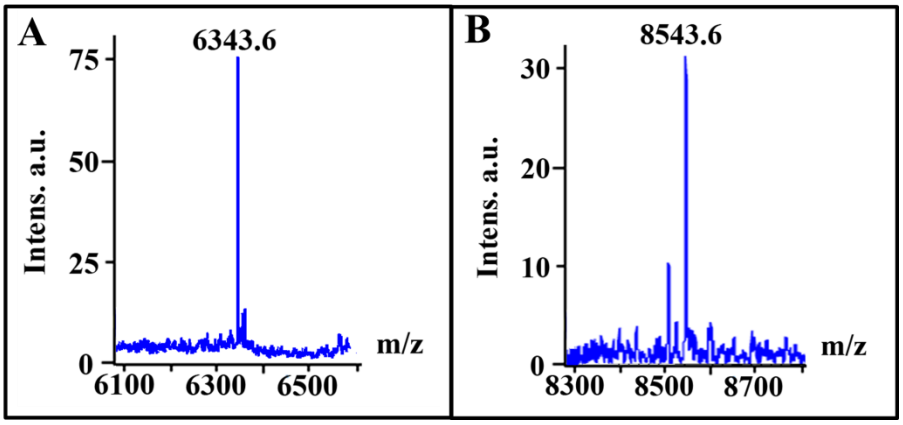
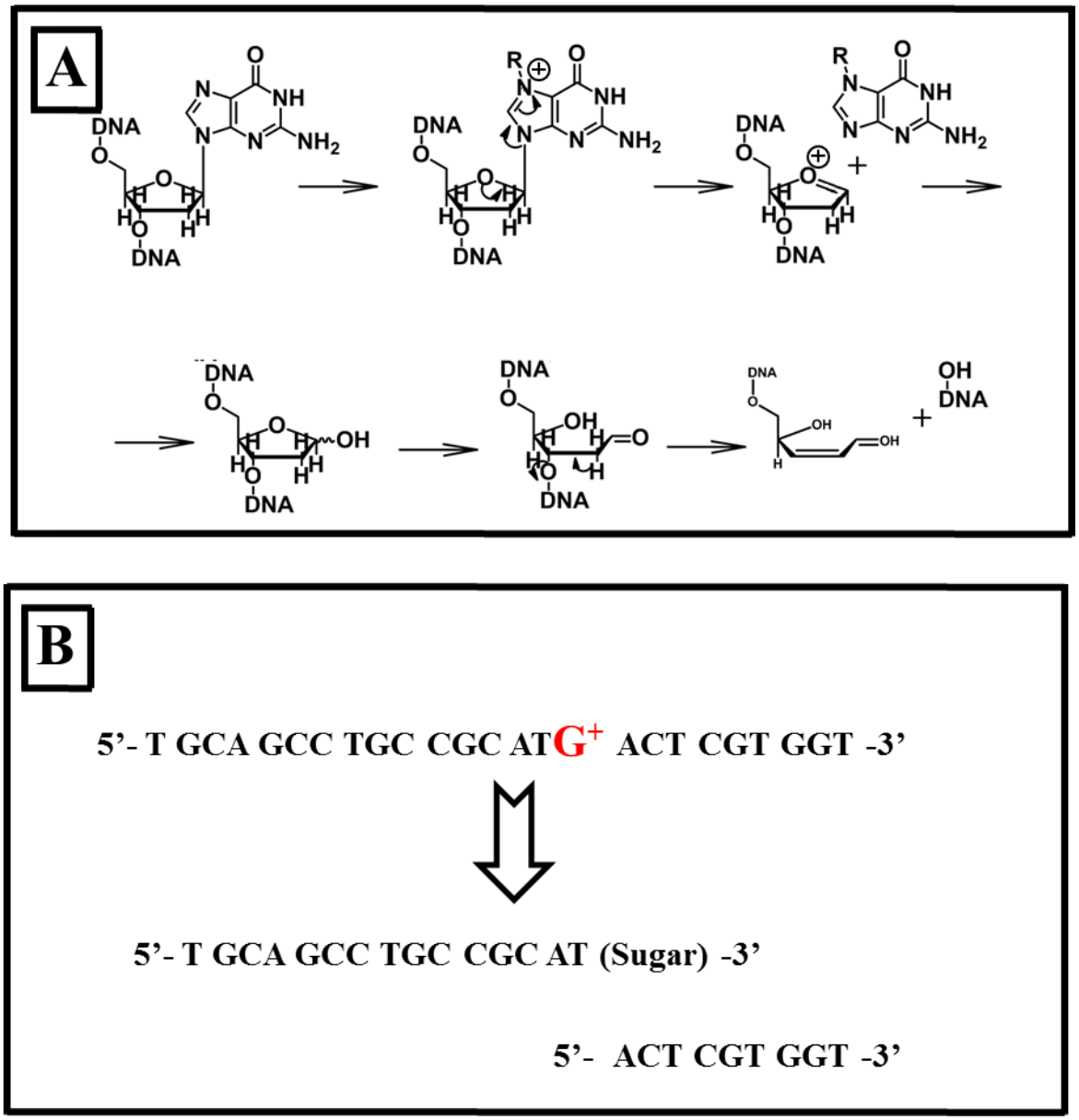


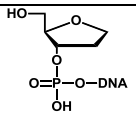
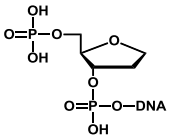
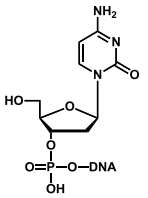
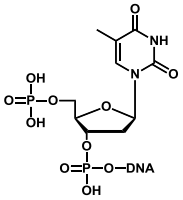
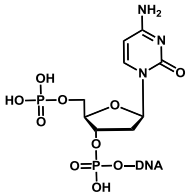
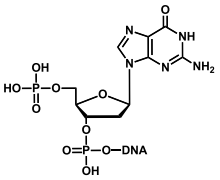
Figure 6.6. MALDI-TOF spectrum of: **A:** Peak eluted at 35.6 min is assigned to sulfo-Cy5-labeled recognition sequence ($M_{observed}$ = 6343.6 Da, M_{calc} = 6346.3 Da); **B:** Peak eluted at 39.9 min is assigned at sulfo-Cy5 labeled target oligonucleotide ($M_{observed}$ = 8543.6 Da, M_{calc} = 8548.4 Da).

Alkylation of the N7 position of the guanine residues in DNA places a formal positive charge on the guanine ring system thus converting the base to a good leaving group for the heteroatom-assisted S_N1 type hydrolysis reaction (**Scheme 6.2. A**). After depurination, the newly formed abasic site transforms to the aldehyde form. Deprotonation of the C2' position in the aldehyde form can lead to DNA strand cleavage *via* elimination of the 3'-phosphate group[24]. After increasing the ionization strength that should facilitate DNA decomposition, mass spectrum peaks consistent with its product of degradation (**Scheme 6.2. B, Fig. 6.7.**).



Scheme 6.2. Depurination of N7-alkylated DNA strand. **A**: Reaction pathway of DNA depurination at neutral pH; **B**: Proposed scheme of depurination of sulfo-Cy5 labeled target DNA[24].

Table 6.1. Proposed species after N7-alkylated DNA degradation.

Entry	3’- sequence-5’	X	$M_{calculated}$
1	3’-TGG TGC TCA X		2908.51
2	3’-TGG TGC TC X		2675.42
3	3’-TGG TGC T X		2415.43
4	3’-TGG TGC X		2206.35
5	3’-TGG TG X		1902.31
6	3’-TGG T X		1614.02

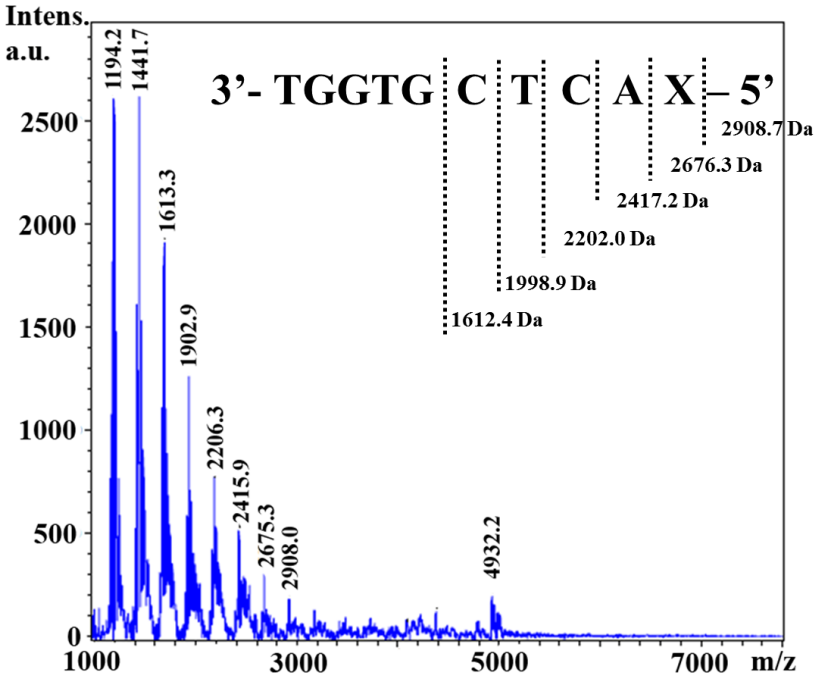


Figure 6.7. MALDI-TOF spectrum of sulfo-Cy5 labeled target oligonucleotide under high laser intensity during ionization.

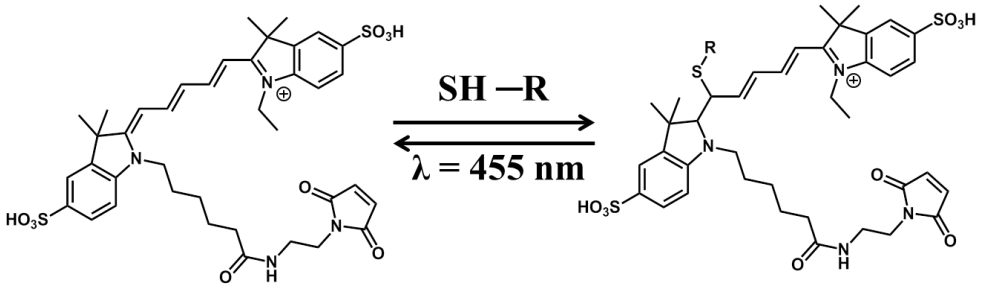
Then, we tested the modified strand in FRET, using the 25-mer complementary DNA strand having the biotin at 5'-end and Cy3 moiety as the donor fluorophore. If the target DNA in the mixture was labeled sufficiently in the site-specific manner, the distance between the two dyes is around 10 nm after annealing. FRET results should be detected by direct excitation of the Cy3-containing reference strand (donor) at λ = 550 nm and observation of an increased Cy5 emission (acceptor). In case of this experiment, only 2% of traces demonstrated energy transfer with increasing emission in the Cy5 region (see experimental part).

6.4 Inactivation of cyanine dyes with nucleophilic or reducing agent.

The Cy3/Cy5 pair is indisputably the most popular donor and acceptor pair for smFRET because (a) their spectral separation is large (~100 nm); (b) they are both photostable in oxygen-free environments; (c) the quantum yields of both dyes are comparable (~0.2); and (d) they are commercially available in amino-, thiol- and carbonyl-reactive forms.

Despite all advantages these fluorophores still possess two main problems for single-molecule biophysical studies using fluorescence, which are photobleaching and undesired intensity fluctuations. The third problem was recently identified by D.R. Liu and concerns the instability of the dye core[25]. After incubation of Cy5 in the presence of the strong nucleophile 2-mercaptoethanol, he detected formation of a new product. The mass spectrum and fragmentation pattern was consistent with thiol attachment to the polymethine bridge of the fluorophore

(**Scheme 6.3**). Fluorescence measurements showed a corresponding loss of fluorescence after red excitation and recovery following UV illumination. Additionally, J.D. Puglisi et al.[26] identified decreasing emission intensity from cyanine dyes after treatment with TCEP due to modification of the polyethene bridge.



Scheme 6.3. Proposed mechanism of Cy5 dye inactivation according to Liu et al[25].

The final step in our approach is based on the quantitative reduction of the disulfide bridge with TCEP and capturing of the newly formed reactive group by maleimide-sulfo-Cy5. Unfortunately, we did not observed any efficient energy transfer after annealing the modified target oligonucleotide with the complementary primer. We assumed that simultaneous presence of the cyanine fluorophore and TCEP during the labeling procedure inactivated the fluoregenic properties of the Cy5 core. To examine this problem we incubated Sulfo-Cy5 dye in the presence of TCEP (**Figure 6.8.**). RP-HPLC chromatogram (λ = 650 nm) shows complete disappearance of the initial Cy5 peak (retention time 32 min) and formation of an intense peak (RT = 30 min). Maleimido-sulfo-Cy3 dye demonstrated similar behavior during incubation with TCEP (**Figure 6.9**).

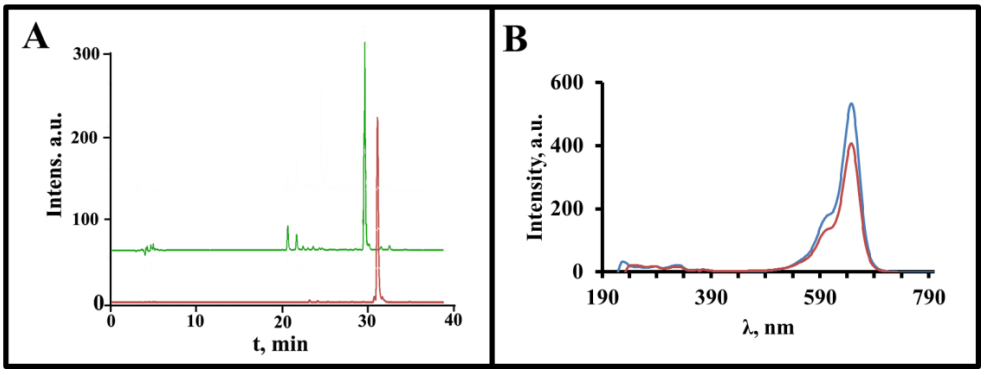


Figure 6.8. **A:** Reverse phase HPLC chromatogram showing instability of maleimide-sulfo-Cy5 dye (λ = 650 nm). Red curve demonstrated chromatogram of initial maleimide-sulfo-Cy5 dye, **2**: maleimide-sulfo-Cy5 after incubation with TCEP for 2 hours, **B:** Absorption spectra of maleimido-sulfo-Cy5 dye before (blue curve) and after incubation with TCEP (red curve).

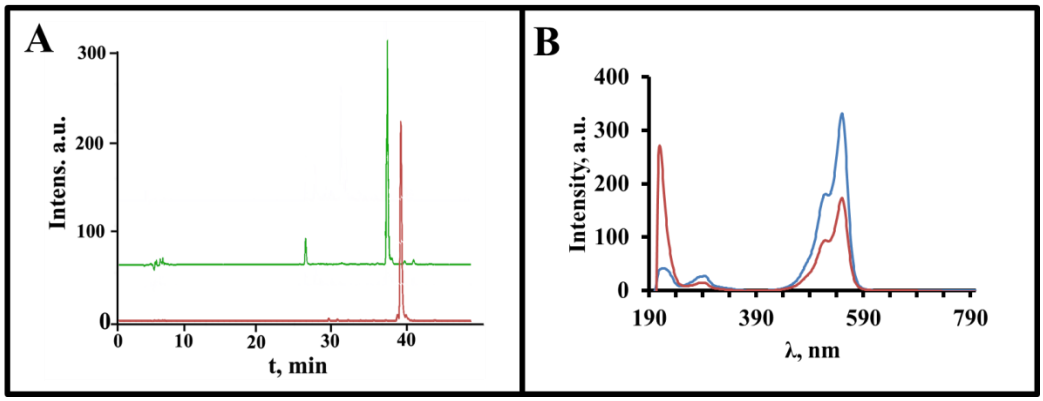


Figure 6.9. **A:** Reverse phase HPLC chromatogram showing instability of maleimide-sulfo-Cy3 dye (λ = 500 nm). Red curve demonstrated chromatogram of initial maleimide-sulfo-Cy3 dye, **2**: maleimide-sulfo-Cy3 after incubation with TCEP for 2 hours, **B:** Absorption spectra of maleimido-sulfo-Cy3 dye before (blue curve) and after incubation with TCEP (red curve).

6.5 Site-specific labeling with donor (sulfo-Cy3) fluorophore

Previous results showed that the sulfo-Cy5 fluorophore is apparently too unstable to be incorporated into DNA. Hence, we tried the labeling with the more stable sulfo-Cy3 dye[23]. In addition we used strands that form a 40 nucleotides long double stranded region after annealing which leads to a two fold increase of the helix stability and hence to reach up (-82.6) kcal/mol. Duplex with low free energy should stabilize the structure sufficiently to arrange donor and acceptor fluorophores with a fixed distance between them. This allows us to expect only two FRET states, i.e. “on” and “off” and to avoid any intermediate states that appear according to “DNA breathing ends”. To test the applicability we incubated the reactive sequence SRS_6.2 containing a cystamine linker with ODN_6.3 containing guanine in the (n+1) position under the usual modification conditions, i.e. 0.1 M MOPS buffer at pH = 7.4, 1 M NaCl at RT. After 24 hours, the disulfide bridge was reduced with TCEP. Subsequently, sulfo-Cy3 was either added directly to the reactive mixture (**Fig. 6.10. A**) or TCEP was removed by size-exclusion gel chromatography prior to sulfo-Cy3 addition (**Fig. 6.10 B**).

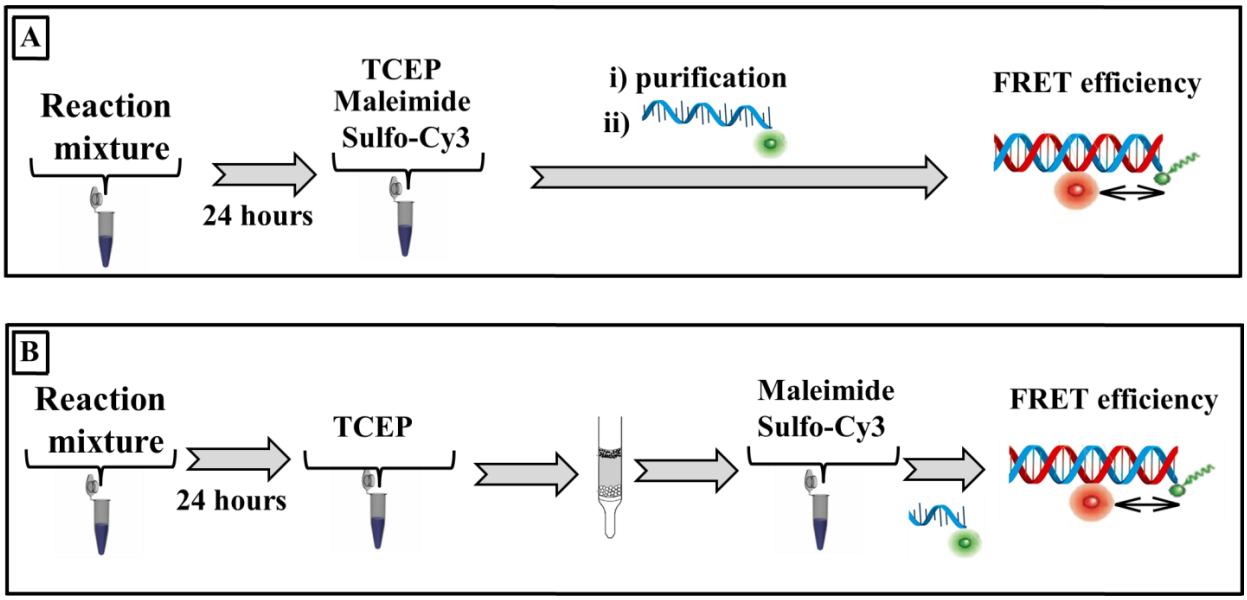


Figure 6.10. Scheme of the strategy for the site-specific labeling of the oligonucleotide. **A**: In the first step, SRS_6.2 was annealed with ODN_6.3 and incubated for 24 hours. Then the disulfide bridge was reduced by TCEP followed by addition of maleimide-sulfo-Cy3 dye. After labeling, oligonucleotide mixture was precipitated by ethanol and used in the fluorescence experiment. **B**: After annealing and reduction of the disulfide bridge, a desalting step to remove TCEP was applied. After it, fluorophore was conjugated to thiol group, followed by purification and fluorescence measurements (excitation λ = 550 nm).

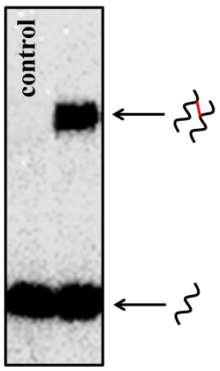


Figure 6.11 Phosphorimage autoradiogram of 12% denaturing PAGE analysis of ICL formation between SRS_6.2 and template ODN_6.3

Following incubation of the two strands at room temperature, we observed the formation of a band with slower migration in the 12% denaturing PAGE gel and yield of 25–30% based on the unmodified strand (**Fig. 6.11**). The cross-linked product was treated with TCEP for 2 hours and desalted by NAP-5 column under nitrogen atmosphere. In the last step, maleimide Sulfo-Cy3 was added in the dark and after 6 hours, the reaction mixture was analyzed by RP-HPLC chromatogram (**Figure 6.12. A**). Elution patterns were monitored with absorption at λ = 260 nm. Similar to the previous results with Sulfo-Cy5 labeling, reverse-phase HPLC chromatogram demonstrated at λ = 260 nm the formation of three major peaks at 13 min, 27 min and 28.5 min. In the case of last two peaks the UV/Vis absorption spectra show clearly the presence of both additional maximum at 560 nm and the strand fluorescent peak with excitation at 550 nm and emission at 570 nm. At the same time peak at 13 min demonstrated UV absorption similar to the unmodified DNAs. These spectroscopic properties lead to the

conclusion that the last two peaks correspond to the sulfo-Cy3-DNA conjugate and the first peak to the unmodified donor or template DNA.

To characterize the products further HPLC was used to isolate the peaks with absorption maximum at λ = 550 nm. According to MALDI-TOF mass spectrometry (**Fig. 6.12. B**), the molecular weight of peak with retention time t = 27 min (6262.0 Da) was in agreement with the calculated values of sulfo-Cy3-labeled reactive sequence (6265.4 Da). For the last peak at t = 29 min the observed molecular weight (13389.2 Da) was found to be consistent with the calculated mass (13390.82 Da) of the sulfo-Cy3-labeled target template.

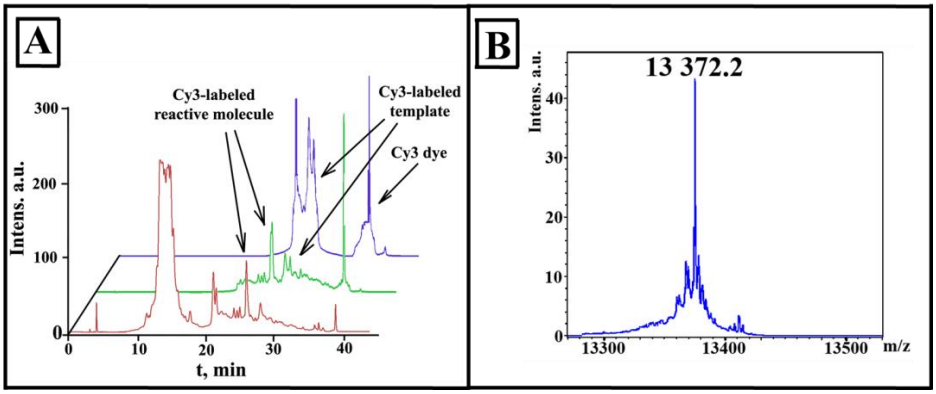


Figure 6.12. Evaluation of oligonucleotide labeling with maleimido-sulfo-Cy3 fluorophore. **A**: RP-HPLC chromatogram of reaction mixture. For correct evaluation we used different detectors with different wavelengths: red curve – λ = 260 nm, green curve – λ = 550 nm, blue curve – fluorescent detector ($\lambda_{em.}$= 550 nm, $\lambda_{ex.}$= 570 nm). **B**: MALDI-TOF spectrum of range from 27 to 30 min that was assigned to sulfo-Cy3-labeled target template ($M_{observed}$= 13372.2 Da, M_{calc}= 13375.4 Da).

After incorporation of the Sulfo-Cy3 fluorophore, the target 40-mer DNA oligonucleotide was annealed with complementary Cy5-containing DNA (ODN_6.4) at a molar ratio of 10:1 in 1 M NaCl in 0.1 M MOPS buffer at pH = 6.9. In the first step we measured the respective emission of single stranded sulfo-Cy3 and Cy5 labeled oligonucleotides at donor excitation wavelength of λ = 550 nm for the correct evaluation of the strand mixture. Excitation of the Cy5 fluorophore at the non-characteristic wavelength of λ = 550 nm instead of λ = 650 nm leads to low emission at λ = 670 nm (**Fig. 6.13 A, green curve**). At the same time emission of Cy3 fluorophore demonstrated a broad shoulder that stretched up to 670 nm (**Fig. 6.13 A, red curve**). Due to the fact that both emission signals of the two non-hybridized strands overlap in the desired region, we used the sum of the two spectra as a baseline for the further measurements. The experimental

emission curves were fitted to the overlaps curves of independent samples and used for normalization of the experimental spectra (**Fig. 6.13**).

The fluorescence spectrum of Cy3-labeled target strand and complementary Cy5-labeled donor sequences after annealing demonstrated an increase of emission at 670 nm that is responsible for Cy5 acceptor dye. To confirm that the increase of fluorescence was observed due to proper incorporation of donor fluorophore and to evaluate input of energy transfer between on increasing emission intensity we extracted baseline Cy3-labeled single strand and final curve (**Fig. 6.13 B red curve**) was compared with Cy5-labeled single strand baseline. Comparison of two spectra demonstrated high intensity in reaction comparable to initial baseline. The emission value of the modified protocol increased 1.5 times from 6 to 10 a.u.

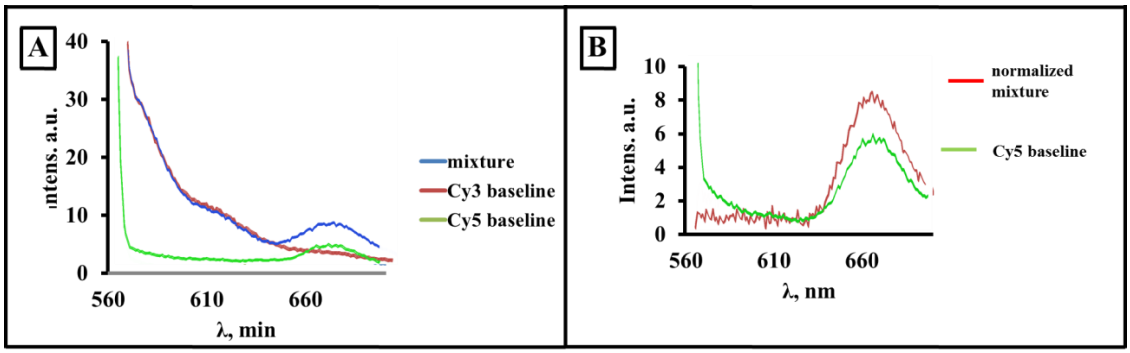


Figure 6.13. Detection by FRET of nucleic acids with a specific sequence (excitation λ = 550 nm). **A**: Changes in fluorescence spectrum caused by hybrid formation. Acceptor probe, 40-mer purchased DNA with Cy5 fluorophore at 5'-end (ODN_6.4). Complementary strand probe was modified by alkylation of guanine at a distance of 10 nucleotides from the donor fluorophore and labeled with sulfo-Cy3 dye in the absence of TCEP. Donor and acceptor strands were mixed at a molar ratio of 10:1 in 1M NaCl in 0.1M MOPS buffer and used as a pair probe. **1**: blue curve - spectrum of the annealed probe. **2**: red curve – sulfo-Cy3-labeled ODN_6.3 that modified by modular system. **3**: green curve – Cy5-labeled ODN_6.4 that was bought form "Lumiprobe Corporation". **B**: normalized fluorescence spectra of hybrids. Green curve - non-target single strand oligonucleotide with aceptor fluorophore. Red curve – emission of reactive mixture with subtraction of baseline signal of donor fluorophore.

In the control experiments (**Fig. 6.14**), in which maleimide-Cy3 and TCEP were added in the same time, after annealing of donor and acceptor DNA sequences we also observed relatively strong increasing in the acceptor region. However, after substracting the Cy3-baseline from reaction emission spectra we observed complete overlapping with Cy5-baseline. We interpreted these result in the way, that no energy transfer was observed, and emission spectra of reactive mixture includes only sum of Cy3- and Cy5- baselines.

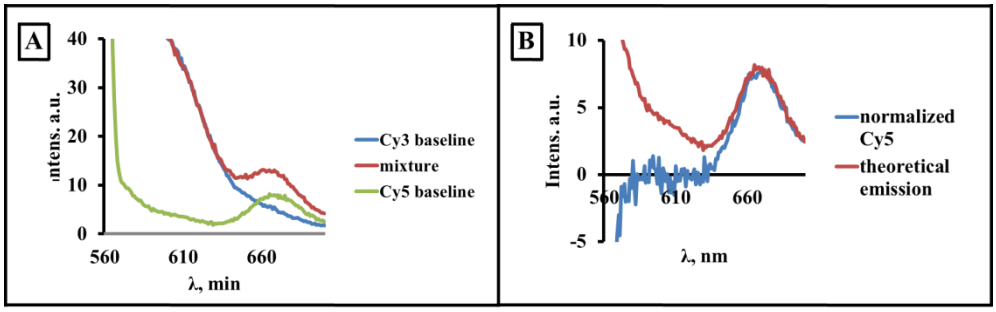


Figure 6.14. Detection by FRET of nucleic acids with a specific sequence (excitation λ = 550 nm). **A**: Changes in fluorescence spectrum caused by hybrid formation. The acceptor probe consists of a 40-mer purchased DNA with Cy5 fluorophore at 5'-end (ODN_6.4). The complementary strand probe was modified by alkylation of guanine in distance of 10 nucleotides from acceptor fluorophore in the acceptor strand and labeled with sulfo-Cy3 dye in the presence of TCEP. Donor and acceptor probes were mixed at a molar ratio of 10:1 in 1 M NaCl in 0.1 M MOPS buffer at pH = 6.9. **1**: red curve - spectrum of the annealed probe. **2**: blue curve – donor probe. **3**: green curve – acceptor probe. **B**: normalized fluorescence spectra of hybrids. Red curve – theoretical emission after overlapping donor and acceptor single stranded probes. Blue curve represents emission of reactive mixture with subtraction of baseline signal of donor fluorophore.

Next, we examined the time course of hybridization between site-specifically modified Cy3-containing oligonucleotide and it complementary Cy5-containing strand (**Fig. 6.15**). Two complementary single stranded DNA were annealed at a molar ratio of 10:1 in 1 M NaCl of 0.1 M MOPS buffer at pH = 6.9, and the fluorescence spectrum was measured at 0 min (red curve), 60 min (green curve), and 180 min (violet curve) after annealing. **Fig. 6.15** shows changes in the fluorescence images during incubation. After 60 min of incubation emission increased in intensity in the Cy5 region from 8 to 10 a.u. and continues to increase for the next 2 hours. Then the fluorescence intensity of the samples decreased and reached initial level of emission in 5 hours after annealing.

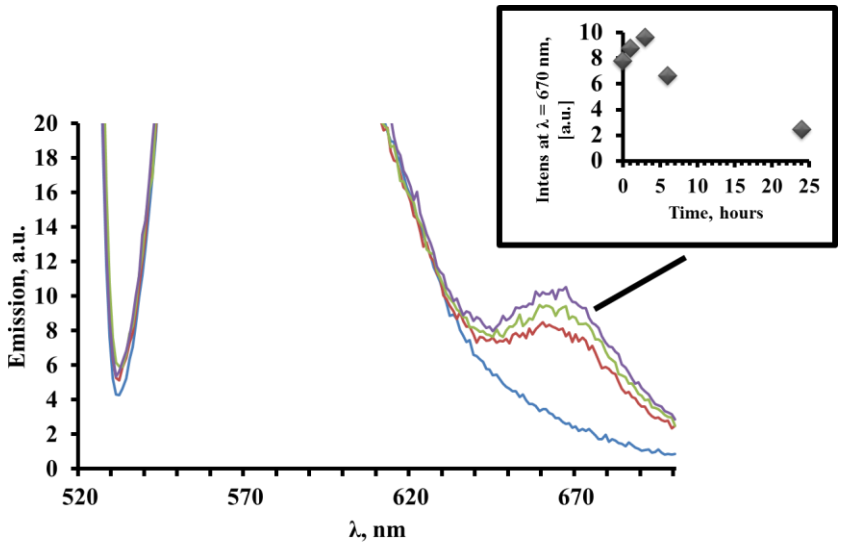


Figure 6.15. Time dependent fluorescence spectral changes by hybrid formation of the sulfo-Cy3/Cy5 oligonucleotide pair (excitation λ = 550 nm). Two complementary single stranded DNA were annealed at a molar ratio of 10:1 in 1 M NaCl of 0.1 M MOPS buffer at pH = 6.9, and the fluorescence spectrum was measured at 0 min (red curve), 60 min (green curve), and 180 min (violet curve) after annealing. Blue curve shows Cy3-containing DNA before addition of the Cy5-labelled strand.

As a control experiment, we mixed acceptor and donor DNA strands in 0.1 M MOPS buffer 1 M NaCl at pH = 6.9 in the 100 µL cuvette and right after mixing measured emission at excitation of 550 nm (**Figure 6.16 A red curve**). This sample was incubated in the same cuvette in the dark for 1 hour and second measuring was done (**Figure 6.16 A blue curve**). To analyze spectral changes during annealing, initial emission spectrum was subtracted from final emission spectrum and the result demonstrates decreasing intensity in the Cy3 region and increasing intensity in the Cy5 region (**Figure 6.16 B**). During 24 hours, the emission dropped continuously, which also indicates that the cyanine fluorophore was dissociated over the time course of the experiment.

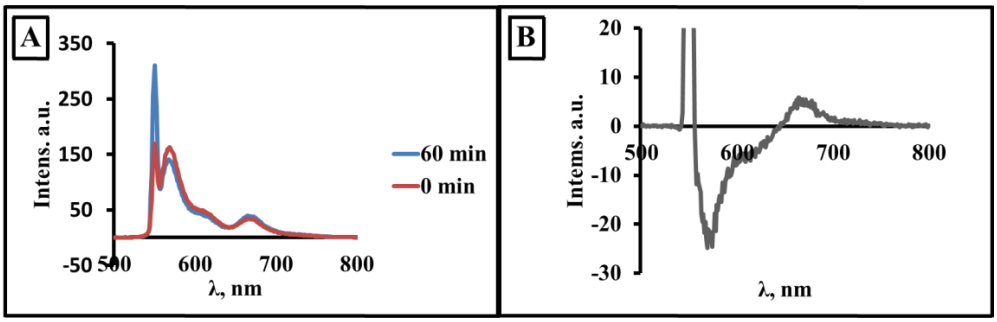


Figure 6.16. Detection by FRET of nucleic acids with a specific sequence (excitation λ = 550 nm). **A**: Changes in fluorescence spectra caused by hybrid formation in the beginning (red curve) and after 60min (blue curve). **B**: Spectral changes during annealing of sulfo-Cy3 and Cy5 labeled complementary DNA primers.

6.6 Conclusion

In this chapter we demonstrated the possible application of the modular system to label long oligonucleotide template with fluorophore. Target template ODN_6.3 contains few guanines and we demonstrated possibility to modify only one of the guanines that is placed at position (n+1relative to reactive group from reactive sequence. After cleavage of the cross-linked product, the newly formed thiol group was labeled with maleimido-derivative of cyanine dyes.

Initial attempts to label thiol-modified target DNA with maleimide-sulfo-Cy5 fluorophore using standard protocol were not successful and no energy transfer between FRET pairs was detected. We sensed the problem in the differences in stability of cyanine fluorophores in presence/absence of reducing agent described above. To modify the labeling protocol we introduced an additional desalting step to remove TCEP. Indeed, after conjugation target template with donor Cy3 dye we observed changes at emission spectra. FRET, which is denoted by a decrease in donor fluorescence and an increase in acceptor fluorescence, was observed in the presence of complementary strand with Cy5 dye acceptor. These results indicate that the site-specific labeling of the modular system shows feasibility for the strategy of site-specific modification of target oligonucleotides.

6.7 References

[1] W. R. Algar, D. E. Prasuhn, M. H. Stewart, T. L. Jennings, J. B. Blanco-Canosa, P. E. Dawson, I. L. Medintz, *Bioconjugate Chem.* **2011**, *22*, 825-858.
[2] K. E. Sapsford, W. R. Algar, L. Berti, K. B. Gemmill, B. J. Casey, E. Oh, M. H. Stewart, I. L. Medintz, *Chem. Rev.* **2013**, *113*, 1904-2074.
[3] Y. Takaoka, A. Ojida, I. Hamachi, *Angew. Chem. Int. Ed.* **2013**, *52*, 4088-4106.
[4] U. Feldkamp, C. M. Niemeyer, *Angew. Chem. Int. Ed.* **2006**, *45*, 1856-1876.
[5] P. Guo, *Methods* **2011**, *54*, 201-203.
[6] D. M. Kolpashchikov, *Chem. Rev.* **2010**, *110*, 4709-4723.

[7] H. Zhang, F. Li, B. Dever, X.-F. Li, X. C. Le, *Chem. Rev.* **2012**, *113*, 2812-2841.
[8] A. N. Glazer, *Annu. Rev. Biochem* **1970**, *39*, 101-130.
[9] G. T. Hermanson, 2-nd ed., Elsevier Science, **2008**, p. 1323.
[10] R. Roy, S. Hohng, T. Ha, *Nat. Methods* **2008**, *5*, 507-516.
[11] T. Ha, P. Tinnefeld, *Annu. Rev. Phys. Chem.* **2012**, *63*, 595-617.
[12] S. Tyagi, F. R. Kramer, *Nat Biotech* **1996**, *14*, 303-308.
[13] P. Zhang, T. Beck, W. Tan, *Angew. Chem.* **2001**, *113*, 416-419.
[14] X.-H. Peng, Z.-H. Cao, J.-T. Xia, G. W. Carlson, M. M. Lewis, W. C. Wood, L. Yang, *Cancer Res.* **2005**, *65*, 1909-1917.
[15] R. Yamamoto, P. K. R. Kumar, *Genes Cells* **2000**, *5*, 389-396.
[16] S. Yunger, L. Rosenfeld, Y. Garini, Y. Shav-Tal, *Nat. Protoc.* **2013**, *8*, 393-408.
[17] L. W. Miller, V. W. Cornish, *Curr. Opin. Chem. Biol.* **2005**, *9*, 56-61.
[18] C. R. Sabanayagam, J. S. Eid, A. Meller, *J. Chem. Phys.* **2005**, *122*, 1103-1106.
[19] R. D. Smiley, T. R. L. Collins, G. G. Hammes, T.-S. Hsieh, *P. Natl. Acad. Sci. USA* **2007**, *104*, 4840-4845.
[20] D. S. Talaga, W. L. Lau, H. Roder, J. Tang, Y. Jia, W. F. DeGrado, R. M. Hochstrasser, *P. Natl. Acad. Sci. USA* **2000**, *97*, 13021-13026.
[21] M. Steiner, K. S. Karunatilaka, R. K. O. Sigel, D. Rueda, *P. Natl. Acad. Sci. USA* **2008**, *105*, 13853-13858.
[22] G. N. Grimm, A. S. Boutorine, C. Helene, *Nucleos. Nucleot. Nucl.* **2000**, *19*, 1943-1965.
[23] M. Levitus, S. Ranjit, *Q. Rev. Biophys.* **2011**, *44*, 123-151.
[24] K. S. Gates, T. Nooner, S. Dutta, *Chem. Res. Toxicol.* **2004**, *17*, 839-856.
[25] G. T. Dempsey, M. Bates, W. E. Kowtoniuk, D. R. Liu, R. Y. Tsien, X. Zhuang, *J. Am. Chem. Soc.* **2009**, *131*, 18192-18193.
[26] C. E. Aitken, R. A. Marshall, J. D. Puglisi, *Biophys. J.* **2008**, *94*, 1826-1835.

CHAPTER 7
Experimental Procedures

7.1 General Methods

Unless otherwise stated, starting materials were obtained in the highest commercial grades and used without further purification. Solvents for organic syntheses were distilled prior to use and if necessary dried under standard conditions. HPLC solvents were obtained in HPLC grade from Roth AG and Sigma Aldrich, deuterated solvents were purchased from Armar Chemicals. Thin layer chromatography (TLC) was performed with Merck TLC Silicagel 60 F254 plates and visualized with UV light, potassium permanganate and Schlittler staining solution. Flash chromatography was carried out using Sigma Aldrich high-purity grade silica gel (pore size 60 Å, 220-440 mesh particle size, 35-75 μm particle size). H_2O used for all manipulations with oligonucleotides was purified with a TKA GenPure dispenser and water as well as buffer solutions for enzymatic reactions were sterilized either by autoclaving or sterile filtration. Unmodified and 3'-phosphorylated oligonucleotides were obtained from Microsynth AG. Enzymes for enzymatic digestion of oligonucleotides, polymerase stop assays and radioactive labeling were purchased from Thermo Scientific (DNase1), USBiological (Phosphodiesterase 1), Promega (Calf Intestinal Alkaline Phosphatase, GoTaq DNA polymerase with 5X Colorless GoTaq Reaction Buffer) and New England Biolabs (T4 Polynucleotide Kinase).

^{1}H and ^{13}C NMR spectra were recorded on a Bruker ARX-300, AV-400 or AV-500 spectrometer. Chemical shifts in ^{1}H and ^{13}C NMR spectra, respectively, were reported in parts per million (ppm) on the δ scale from an internal standard of residual chloroform, methanol or dimethylsulfoxide. Data for ^{1}H NMR were reported as follows: chemical shift, multiplicity (s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet), coupling constant in Herts (Hz) and integration. ESI mass spectra were recorded on a Bruker Esquire 6000 spectrometer and LC-MS measurements were performed using a Macherey-Nagel Nucleosil C18 column (5 μm, 3.0 x 250 mm). MALDI-TOF analyses were performed on an Autoflex time-of-flight mass spectrometer with a 337 nm nitrogen laser. All MALDI-MS spectra were recorded in negative ion mode using a 3-HPA matrix and calculated masses correspond to the molecular weight deducting a proton. Reactions with oligonucleotides were performed on an Eppendorf Thermomixer Comfort and oligonucleotide concentrations were determined using a Thermo Scientific Nanodrop 2000 Spectrophotometer. Elemental microanalyses were performed on a LecoCHNS 932 elemental analyser. Fluorescence measurements were carried out on a Varian Cary Eclipse Fluorescence Spectrophotometer using the software Cary Eclipse Scan Application in combination with a thermostat Varian Cary Water Peltier System PCV 150 to maintain the temperature at 25°C. For

HPLC two different systems were used: Analysis of the model reaction was done on a Hitachi LaChrome ELITE (*HPLC system 1*) and the evaluation of oligonucleotides and digestion mixtures on a Dionex UltiMate 3000 HPLC using the software Chromeleon 7.1 (*HPLC system, 2*). Visualization and quantification of radioactive gels were conducted on a Typhoon Scanner FLA 9500 with ImageQuant TL 1D (GE Healthcare).

Used HPLC methods: **HPLC method 1:** HPLC system 1. Macherey-Nagel Nucleodur C18 Gravity column (5 µm, 4.0 x 250 mm); Solvent "A": CH_3COONH_4 25 mM, pH = 6.0. Solvent "B": methanol. Speed: 1.0 mL/min. Solvent gradient: from 0 min to 5 min "A" = 100%; from 5 min to 20 min – "A" = 100% to 90%;

HPLC method 2: HPLC system 2. Waters X-Bridge C18 column (3.5 µm, 4.6 x 150 mm). Solvent "A": CH_3COONH_4 25 mM, pH = 6.0. Solvent "B": methanol. Speed: 0.5 mL/min. Solvent gradient: from 0 min to 30 min "B" from 0% to 30%.

HPLC method 3: HPLC system 1: Macherey-Nagel Nucleodur C18 Gravity column (5 µm, 4.0 x 250 mm);. Solvent "A": TEAA 100 mM with 1% CH_3CN, pH = 7.0. Solvent "B": Acetonitrile. Speed: 1.0 mL/min. Solvent gradient: from 0 min to 15 min "B" from 0% to 10%; from 15 min to 50 min – solvent "B" from 10% to 60%.

HPLC method 4: HPLC system 2: Macherey-Nagel Nucleodur C18 Gravity column (5 µm, 4.0 x 250 mm). Solvent "A": TEAA 100 mM with 1% CH_3CN, pH = 7.0. Solvent "B": Acetonitrile. Speed: 1.0 mL/min. Solvent gradient: from 0 min to 40 min "B" from 1% to 20%; from 40 min to 80 min – solvent "B" from 20% to 80%.

HPLC method 5: HPLC system 2: Waters X-Bridge C18 column (3.5 µm, 4.6 x 150 mm). Solvent "A": TEAA 100 mM with 1% CH_3CN, pH = 7.0. Solvent "B": Acetonitrile. Speed: 0.5 mL/min. Solvent gradient: from 0 min to 10 min "B" from 5% to 10%; from 10 min to 30 min – solvent "B" from 10% to 20%, from 30 min to 45 min – from 20% to 80%.

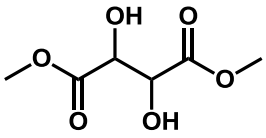
HPLC method 6: HPLC system 2. Waters X-Bridge C18 column (3.5 µm, 4.6 x 150 mm). Solvent "A": TEAA 100mM with 1% CH_3CN, pH = 7.0. Solvent "B": Acetonitrile. Speed: 0.5 mL/min. Solvent gradient: from 0 min to 20 min "B" from 0% to 5%; from 20 min to 30 min – solvent "B" from 5% to 20%, from 30 min to 35 min – from 20% to 80%.

HPLC method 7: HPLC system 2. Waters X-Bridge C18 column (3.5 µm, 4.6 x 150 mm). Solvent "A": TEAA 100 mM with 1% CH_3CN, pH = 7.0. Solvent "B": Acetonitrile. Speed: 0.5 mL/min. Solvent gradient: from 0 min to 20 min "B" from 0% to 5%; from 20 min to 40 min – solvent "B" from 5% to 25%, from 40 min to 50 min – from 25% to 80%.

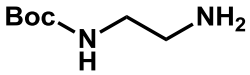
HPLC method 8: HPLC system 2. Waters X-Bridge C18 column (3.5 µm, 4.6 x 150 mm). Solvent "A": TEAA 100 mM with 1% CH_3CN, pH = 7.0. Solvent "B": Acetonitrile. Speed: 1.0 mL/min. Solvent gradient: from 0 min to 50 min "B" from 1% to 40%.

HPLC method 9: HPLC system 2. Waters X-Bridge C18 column (3.5 μm, 4.6 x 150 mm). Solvent "A": TEAA 100 mM with 1% CH_3CN, pH = 7.0. Solvent "B": Acetonitrile. Speed: 0.5 mL/min. Solvent gradient: from 0 min to 40 min "B" from 5% to 30%; from 40 min to 50 min – solvent "B" from 30% to 50%, from 50 min to 80 min – from 50% to 80%.

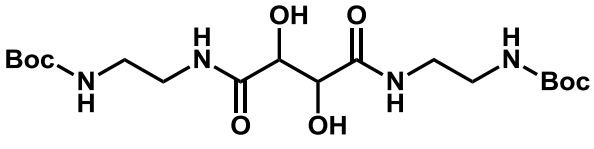
7.2 Synthetic methods



2,3-Dihydroxy-succinic acid dimethyl ester: To a stirring solution of (+/-)-tartaric acid (2.06 g, 13.5 mmol) in methanol (30 mL) was added boric acid (90 mg, 0.14 mmol) in one portion and the mixture soon became homogeneous. The reaction was stirred at room temperature for 18 hours and the bulk solvent was removed *in vacuo* to afford tartrate dimethyl ester as clear oil. The residue was dissolved in Et_2O, filtered through a short plug of Celite, and concentrated to afford 2.2 g (98%) of methyl (+/-)-tartrate. The ^{1}H-NMR chemical shifts matched values for authentic samples of the dimethyl ester. ^{1}H NMR (400 MHz, CD_3OD): δ 4.82 (s, 2 H), 3.98 (s, 6 H), ESI-MS (m/z): $[M+Na]^+$ calcd. for $C_6H_{10}O_6$, 202.1; found, 202.3.

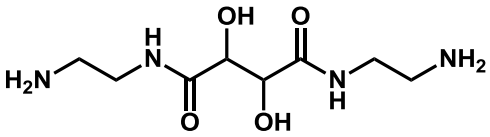


1,2-Diamino-N-tert-butyloxycarbonylethane: 22 mmol of 1,2-diaminoethane was stirred in 100 mL of $CHCl_3$. 960 mg of $(Boc)_2O$ (4.4 mmol) dissolved in 50 mL of $CHCl_3$ was added dropwise at 0°C. The reaction mixture was stirred for 2 h at room temperature. The solution was filtered and the filtrate was concentrated. The oily residue was diluted with AcOEt and washed with brine. The aqueous layer was extracted with AcOEt. The organic layers were dried over $MgSO_4$ and the solvent was evaporated. 600 mg of compound was obtained as a colourless oil (85%). ^{1}H NMR ($CDCl_3$, 300MHz): δ 1.38 (s, 9 H), 1.60 (m, 2 H), 2.75 (t, 2 H, J = 6.5 Hz), 3.09 (q, 2 H, J = 6.5 Hz), 5.13 (m, 1 H). ESI-MS (m/z): $[M+H]^+$ calcd. for $C_7H_{16}N_2O_2$, 161.1; found, 161.0.

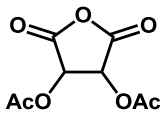


*N,N'-Di-(2-tert-butoxycarbonylaminoethyl)-L-tartaramide***:** 2,3-Dihydroxy-succinic acid dimethyl ester (1.27 g, 7.13 mmol) was added to a solution of mono-*N*-Boc-ethylenediamine (2.40 g, 15.0 mmol) and triethylamine (0.5 mL) in

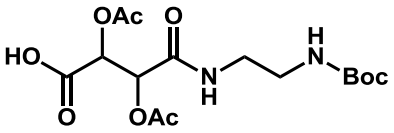
MeOH (35 mL). The solution was heated at reflux for 20 h under an nitrogen. After cooling to room temperature, the solution was concentrated under reduced pressure. The pale yellow residue was treated with 1:19 MeOH-Et_2O (30 mL), which produced a white solid and a yellow solution. The solid was filtered, washed with Et_2O (30 mL), and dried under vacuum to 2.08 g (67%). ^{1}H NMR (400 MHz, DMSO-*d*6): δ 7.82 (t, 2 H, *J* = 5.6 Hz), 6.85 (t, 2 H, *J* = 5.2 Hz), 5.47 (d, 2 H, *J* = 7 Hz), 4.21 (d, 2 H, *J* = 7 Hz), 3.14 (m, 4H), 3.00 (q, 4 H, *J* = 6 Hz), 1.38 (s, 18 H); ^{13}C NMR (400 MHz, DMSO-d6): δ 172.1, 155.6, 77.6, 72.5, 38.6, 28.2; ESI-MS (m/z): $[M+Na]^+$ calcd. for $C_{18}H_{34}N_4O_8$, 457.2; found, 457.1.



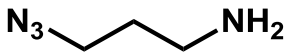
*N,N'-Di-(2-aminoethyl)-tartaramide***:** N,N-Di-(2-tert-butoxycarbonylaminoethyl)-tartaramide (1.00 g, 2.30 mmol) was treated with anhydrous 4.0 M HCl in 1,4-dioxane (10 mL, 40 mmol), and the resulting suspension was stirred under an N_2 atmosphere at room temperature for 1 h. The mixture was concentrated under reduced pressure to give a white powder (0.91 g after drying under vacuum) that was purified by ion exchange chromatography using Dowex 2 x8 20 – 50 mesh resin. The column was eluted with H_2O followed by increasing concentrations of aqueous NH_4HCO_3 (0.10 M, 0.15 M, 0.20 M, 0.30 M), and the appropriate fractions were collected and concentrated under reduced pressure. The white residue was dissolved with H_2O and lyophilized to dryness to provide 0.5 g (97%). ^{1}H NMR (400 MHz, D_2O, 0.1 M TFA): δ 4.61 (s, 2 H), 3.59 (m. 4 H), 3.19 (m, 4 H); ^{13}C NMR (400 MHz in D2O, 0.1 M TFA): δ 175.3, 73.0, 39.7, 37.3. ESI-MS (m/z): $[M+H]^+$ calcd. for $C_8H_{18}N_4O_4$, 235.14; found, 235.2.



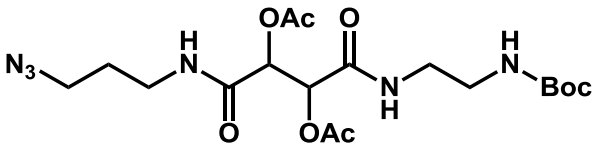
O,O'-Diacetyl-tartaric anhydride: A suspension of tartaric acid (10.00 g, 63 mmol) in acetic anhydride (25 mL) was treated with concentrated sulfuric acid (0.5 mL) to produce an exothermic reaction. The resulting homogeneous solution was refluxed for 10 min and then cooled to room temperature. Further cooling with an ice water bath produced a white crystalline solid that was filtered and washed with Et_2O (3 x 10 mL). The white crystals were then suspended in Et_2O (30 mL) and stirred in an ice water bath for 10 min. After filtering and washing with Et_2O (3 x 10 mL), the crystals were dried under vacuum to 11.93 g (81%). ^{1}H NMR (400 MHz, $CDCl_3$): δ 5.69 (s, 2 H), 2.24 (s, 6H); ^{13}C NMR (400 MHz, $CDCl_3$): δ 170.0, 163.6, 72.2, 20.2. ESI-MS (m/z): $[M+H]^+$ calcd. for $C_8H_8O_7$, 217.0; found, 217.1, [M (methyl ester)+Na]$^+$ calcd. for $C_9H_{12}O_8$, 271.2; found, 271.2.



2,3-Diacetoxy-N-(2-tert-butoxycarbonylamino-ethyl)-succinamic acid: O,O'-Diacetyl-tartaric anhydride (2.00 g, 9.2 mmol) was dissolved in anhydrous THF (135 mL) and then treated dropwise with 1,2-diamino-N-tert-butyloxycarbonylethane (7.3 g, 46 mmol) over 30 min. The mixture was stirred at room temperature for 18 h. The resulting white precipitate was filtered, washed with THF, and then partitioned between EtOAc (50 mL) and 1.0 M HCl saturated with NaCl (50 mL). The layers were separated and the aqueous layer extracted with additional EtOAc (3 x 50 mL). The combined EtOAc layers were dried over Na_2SO_4, filtered, and concentrated under reduced pressure to a white solid. The solid was washed with Et_2O (25 mL) and then dried under vacuum to 1.7 g (75%). ^{1}H NMR (400 MHz, DMSO-d6): δ 8.16 (t, 1 H, J = 5.8 Hz), 8.06 (q, 1 H, J = 4.6 Hz), 6.78 (t, 1 H, J = 6 Hz), 5.44 (d, 1 H, J = 2.6 Hz), 5.42 (d, 1 H, J = 2.6 Hz), 3.11 (m, 1 H), 3.03 (m, 1 H), 2.93 (q, 2 H, J = 6 Hz), 2.08 (s, 6H), 1.37 (s, 9H,) ^{13}C NMR (400 MHz, DMSO-d6): δ 169.4, 169.3, 166.1, 166.0, 155.6, 77.8, 38.8, 28.2, 25.6, 20.5; ESI-MS (m/z): $[M+H]^+$ calcd. for $C_{15}H_{24}N_2O_9$, 377.1; found, 377.2.

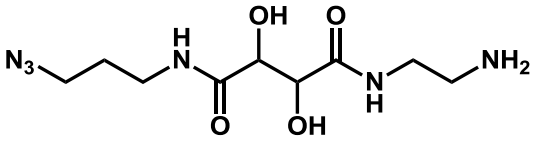


3-Azido-propylamine: Solution of 1-bromo-3-aminopropane hydrobromide (5.47 g, 25.0 mmol, 1.0 equiv) and sodium azide (3.25 g, 50.0 mmol) in 20 mL water was heated at 80 °C for 24 h. The reaction mixture was cooled in an ice bath followed by addition of diethyl ether (30 mL). The pH was adjusted to 14 by addition of KOH pellets keeping the temperature below 10 °C. After separation of the organic phase the aqueous was further extracted with diethyl ether. The combined organic layers were dried over $MgSO_4$ and concentrated in vacuo that resulted in 21.0 mmol (84%) Compound was used without further purification. ^{1}H NMR (400 MHz, CCl_3D): δ 4.72 (t, 2 H, J = 5.6 Hz), 2.60 (t, 2 H, J = 1.2 Hz), 1.95 (m, 2 H).

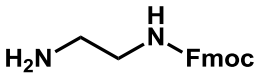


Acetic acid 2-acetoxy-2-(3-azido-propylcarbamoyl)-1-(2-tert-butoxycarbonylamino-ethylcarbamoyl)-ethyl ester: A solution of *N,N'*-dicyclohexylcarbodiimide (3.34 g, 16.2 mmol) in CH_2Cl_2 (60 mL) was cooled to 0 °C and then treated with 2,3-Diacetoxy-N-(2-tert-butoxycarbonylamino-ethyl)-succinamic acid (6.12 g, 16.2 mmol). After stirring the resulting white suspension at 0 °C for 30

min, a solution of 3-azido-propylamine (1.62 g, 16.2 mmol) in CH_2Cl_2 (20 mL) was added dropwise. The resulting mixture was brought to room temperature and stirred under nitrogen for 20 h. The *N,N'*-dicyclohexylurea by-product was filtered and washed with CH_2Cl_2. Concentration of the filtrate produced a yellow residue that was crystalized from 1:4 MeOH-Et_2O (50 mL) to give a white. The white solid was dried under vacuum to 2.92 g. (47%). ^{1}H NMR (400 MHz, CD_3OD): δ 5.72 (d, 2 H, J = 1.2 Hz), 4.76 (t, 2 H, J = 8.2 Hz), 3.51 (t, 2 H, J = 1.2 Hz), 3.43 (t, 2 H, J = 1.6 Hz), 3.29 (t, 2 H, J = 1.2 Hz), 2.19 (m, 2 H), 2.07 (s, 6 H), 1.41 (s, 9 H). ESI-MS (m/z): $[M+Na]^+$ calcd. for $C_{18}H_{30}N_6O_8$, 481.3; found, 482.2.

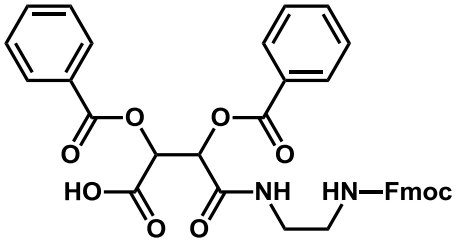


N-(2-Amino-ethyl)-N'-(3-azido-propyl)-2,3-dihydroxy-succinamide: A solution of acetic acid 2-acetoxy-2-(3-azido-propylcarbamoyl)-1-(2-tert-butoxycarbonylamino-ethylcarbamoyl)-ethyl ester (1.75 g, 3.8 mmol) in 1:9 MeOH-CH_2Cl_2 (20 mL) was treated with *p*-toluenesulfonic acid monohydrate (1.96 g, 11.4 mmol). After stirring at room temperature for 24 h, the white precipitate was filtered and washed with 1:9 MeOH-CH_2Cl_2 (2 x 5 mL). The white solid was purified by ion exchange chromatography using Dowex 2 x8 20 – 50 mesh resin. The column was eluted with H_2O followed by 0.30 M NH_4HCO_3 and the appropriate fractions were collected and concentrated under reduced pressure. The white residue was dissolved with H_2O and lyophilized to dryness to provide 0.95 g (91%) of a white foam-like solid. ^{1}H NMR (400 MHz, CD_3OD): δ 4.66 (t, 2 H, J = 5.6 Hz), 4.54 (s, 2 H), 3.43 (t, 2 H, J = 6.8 Hz), 3.20 (t, 2 H, J = 1.8 Hz), 2.82 (d, 2 H, J = 1.6 Hz), 2.11 (m, 3 H). ESI-MS (m/z): $[M+Na]^+$ calcd. for $C_9H_{18}N_6O_4$, 297.1; found, 297.2.

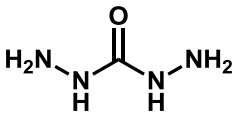


(9H-fluoren-9-yl)-methyl-3-aminopropylcarbamate: 1,2-Diamino-N-tert-butyloxycarbonylethane (2 g, 12.4 mmol) was added to a mixture of 20 mL of CH_2Cl_2 and 20 mL of aqueous saturated $NaHCO_3$. 9-Fluorenylmethylchloroformate (Fmoc-Cl) (3.2 g, 12.4 mmol) was then added to the reaction mixture as a solid portion. The mixture was stirred at room temperature overnight. The water layer was then removed, and the organic layer was washed with water three times, dried over Na_2SO_4, concentrated and purified with flash column chromatography ($CHCl_3$:MeOH 98:2). The relevant fractions were combined, and the solvents were removed under vacuum, affording the *N*-Fmoc-*N*'-Boc protected diaminoethyl (3.78 g, 82%). The Boc group was then quantitatively removed by treatment with 10 mL of 1:1 TFA:CH_2Cl_2 for 2 hours at RT. The final product was obtained by removal of solvents *in vacuo* and recrystallization from 1:1 hexane:ethyl acetate (2.65 g, 76%). ^{1}H NMR (400 MHz, CCl_3D): δ

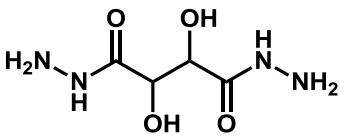
7.92 (d, 2 H, J = 5.6 Hz), 7.80 (m, 2 H), 7.59 (t, 2 H, J = 6.8 Hz), 7.56 (d, 2 H, J = 5.6 Hz), 4.62 (d, 2 H, J = 6.2 Hz), 4.48 (t, 1 H, J = 5.6 Hz), 3.47 (t, 2 H, J = 2.6 Hz), 2.93 (t, 2 H, J = 1.8 Hz). $[M+Na]^+$ calcd. for $C_{17}H_{18}N_2O_2$, 305.1; found, 305.0



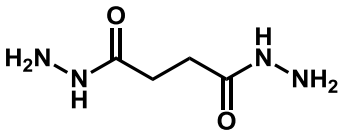
2,3-Dibenzoyloxy-N-(9H-fluoren-9-yl)-methyl-3-aminopropylcarbamate)-succinamic acid: A solution of the *(9H-fluoren-9-yl)-methyl-3-aminopropylcarbamate* (0.84 g, 3 mmol) in dry toluene (10 mL) was stirred magnetically at 0 °C. The 2,3-dibenzoyloxysuccinic acid (0.36 g, 1 mmol) was added in one lot. When the slightly exothermic reaction had subsided (2 min), the ice-bath was removed. The mixture was stirred at room for 6 h. The toluene solution was diluted with 1 M Na_2CO_3 (10 mL). The organic fraction was additionally extracted with water (25 mL) and the combined aqueous solutions were acidified with 5 M HCl. The resulting precipitate was taken into ether and the aqueous phase was extracted with additional ether. The combined organic extracts were washed with brine and dried ($MgSO_4$). The final product was obtained by removal of solvents *in* (0.4 g, 65%). ^{1}H NMR (400 MHz, DMSO-d6): δ 8.02 (d, 1 H, J = 8.4 Hz), 7.86 (d, 2 H, J = 4.6 Hz), 7.69 (t, 2 H, J = 6 Hz), 7.63 (d, 2 H, J = 5.6 Hz), 7.57 (t, 2 H, J = 2.6 Hz), 7.53 (d, 2 H, J = 6.8), 7.47 (m, 2 H), 7.41 (m, 2 H), 5.86 (s, 2 H), 4.60 (t, 1 H, J = 6.2 Hz), 4.48 (d, 2 H, J = 5.6 Hz), 3.47 (t, 2 H, J = 2.6 Hz), 2.93 (t, 2 H, J = 2.6 Hz). ESI-MS (m/z): $[M+Na]^+$ calcd. for $C_{35}H_{30}N_2O_9$, 645,2; found, 645.8.



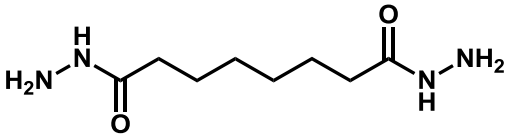
Carbohydrazide[1]: The mixture of dimethyl carbonate (0.09 mol) and hydrazine hydrate (0.10 mol) was stirred at 70 °C for 4 h. The reaction was monitored by TLC. Then the liberated methanol, water, and excess of DMC were removed under reduced pressure. To the residue, additional hydrazine hydrate (0.22 mol) was added. The mixture was stirred at 70 °C for another 4 h. Then the resulting solution was distilled under reduced pressure until the white solid appeared; the residue was cooled slowly and the formed crystal was collected by filtration and dried to give product. Yield: 86%.



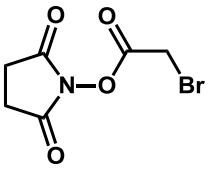
10 mL of hydrazine hydrate were added to solution of 2,3-dihydroxy-succinic acid dimethyl ester (1.00 g, 6.5 mmol) in 30 mL of methanol. The solution was heated at reflux for 40 h under an N_2 atmosphere. After cooling to room temperature, white precipitate was formed. ^{1}H NMR (400 MHz, CD_3OD): δ 4.92 (s, 2 H). ESI-MS (m/z): $[M+H]^+$ calcd. for $C_4H_{10}N_4O_4$, 179.1; found, 179.2. $[M+Na]^+$, 202,3; found, 202.4.



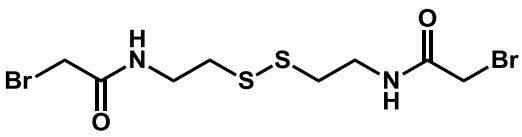
10 mL of hydrazine hydrate were added to solution of dimethyl succinate (0.94 g, 6.5 mmol) in 30 mL of methanol. The solution was heated at reflux for 40 h under an N_2 atmosphere. After cooling to room temperature, white precipitate was formed. ^{1}H NMR (400 MHz, CD_3OD): δ 2.84 (s, 4 H) ESI-MS (m/z): $[M+H]^+$ calcd. for $C_4H_{10}N_4O_2$, 146.1; found, 146.1. $[M+Na]^+$, 169,1; found, 169.1.



10 mL of hydrazine hydrate were added to solution of dimethyl succinate (0.94 g, 6.5 mmol) in 30 mL of methanol. The solution was heated at reflux for 40 h under an N_2 atmosphere. After cooling to room temperature, white precipitate was formed. ^{1}H NMR (400 MHz, DMSO-d6): δ 2.16 (t, 4 H, J = 1.6 Hz), 1.54 (m, 4 H), 1.21 (t, 4 H, J = 1.2 Hz). ESI-MS (m/z): $[M+H]^+$ calcd. for $C_8H_{18}N_4O_2$, 203.1; found, 203.0. $[M+Na]^+$, 225.9; found, 226.0.

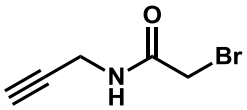


N-Hydroxysuccinimidyl 2-Bromoacetate: To a solution of *N*-hydroxysuccinimide (0.6 g, 5.4 mmol) and 2-bromoacetic acid (0.75 g, 5.4 mmol) in dioxane (30 mL) was added *N*,*N*-dicyclohexylcarbodiimide (1.3 g, 6.5 mmol). The resulting white suspension was stirred for 2 h and then filtered into 100 mL of hexan. The resulting white precipitate was collected by filtration, washed with petroleum ether, and dried *in vacuo* to provide 0.5 g (38%) of *N*-hydroxysuccinimidyl 2-bromoacetate as a white, crystalline solid: ^{1}H NMR (DMSO-d_6) δ 2.82 (s, 4H), 4.62 (s, 2H); ^{13}C NMR (DMSO-d_6) δ 23.1, 26.0, 164.7, 170.3; ESI-MS (m/z): $[M+H]^+$ calcd. for $C_6H_6BrNO_4$ 234.9, found, 234.9; calcd. for $C_3H_5BrO_2$ (Methyl-2-bromoacetate, product of Methanol incorporation) $[M+H]^+$ 152.9, found, 152,9.

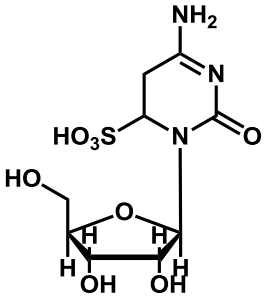


2-Bromo-N-(2-[2-(2-bromo-acetylamino)-ethyldisulfanyl]-ethyl)-acetamide: A solution of cystamine (0.2 g, 1.3 mmol) in water (100 mL) was cooled to 0 °C in ice-water bath. Then the solution of N-hydroxysuccinimidyl 2-bromoacetate (3.1 g, 13 mmol) in CCl_3H (20 mL) and NaOH (10 mmol) was added dropwise. The mixture was

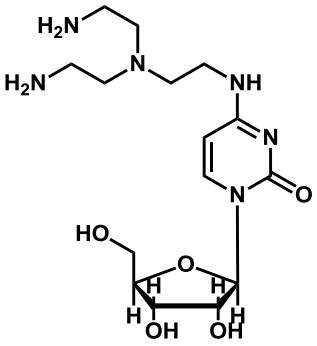
stirring for 2 hours at RT. After that mixture was extracted with CCl_3H (3 x 50 mL). Organic phase was dried over $MgSO_4$ and evaporated. Yield (74%). ^1H NMR (DMSO-d_6, 300 MHz): δ 7.3 (d, 2 H, J = 7.9 Hz); 7.0 (d, 2 H, J = 7.9 Hz); 5.1 (s, 2 H); 3.42 (broad s, 2 H); 2.08 (s, 2H). ESI-MS (m/z): $[M+Na]^+$ calcd. for $C_8H_{14}Br_2N_2O_2S_2$, 415.4; found, 415.3.



2-bromo-N-(propargyl)acetamide: Dry $CHCl_3$ (60 mL) and bromoacetylchloride (1.50 mL, 18 mmol) were added to an oven dried 250 mL round bottom flask under N_2, and cooled at water – ice bath. To this, a solution of propargyl amine (1.22 mL, 18 mmol) and DIEA (3.2 mL, 18 mmol) in dry $CHCl_3$ (20 mL) was added dropwise. Once addition was complete, the reaction was allowed to stir for an additional 1 hour. The reaction mixture was then concentrated under reduced pressure, resuspended in ethyl acetate, and filtered through a short silica plug. The crude material was concentrated and recrystallized from ethyl acetate/hexanes to afford the bromoacetamide as a white solid. 58% yield. ^1H NMR ($CDCl_3$, 400 MHz) δ 2.32 (t, J = 24 Hz, 1H), 4.05 (s, 2H), 4.09 (dd, J = 2.7, 6 Hz, 2H), 6.71 (br, 1H); ESI-MS (m/z): $[M+Na]^+$ calcd. for C_5H_6BrNO, 175.9; found, 175.9.



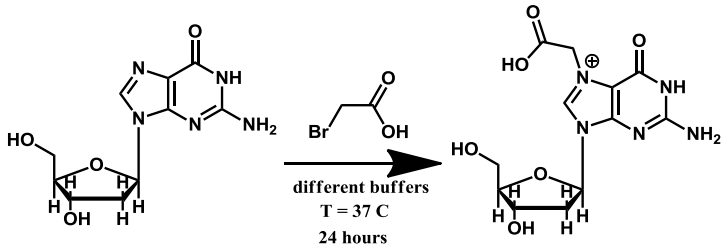
6-amino-3-(4-hydroxy-5-hydroxymethyl-tetrahydro-furan-2-yl)-2-oxo-2,3,4,5-tetrahydro-pyrimidine-4-sulfonic acid: Cytosine (0.5 g, 2.0 mmol) was dissolved in 1M sodium bisulfite sulition at pH =5.6. The solution was mixed at 37°C for 30 min. The solution was then filtered and concentrated *in vacuo*. Purification by flash chromatography (DCM:MeOH, 95:5) and recrystallization from EtOH yielded white crystals of **4b** (287 mg, 1.5 mmol, 83% yield). ESI-MS (m/z): $[M-H]^-$ calcd. for $C_9H_{15}N_3O_8S$, 326.0; found, 326.1.



4-(([bis-(2-amino-ethyl)-amino]-methyl)-amino)-1-(4-hydroxy-5-hydroxymethyl-tetrahydro-furan-2-yl)-1H-pyrimidin-2-one: Tris(2-aminoethyl)amine (0.22 g, 1.5 mmol) was added to solution of 6-sulfo-cytosine (0.05 g, 0.15 mmol) and pH was adjusted to 5.6. After incubation at 37 °C for 2 hours, pH was increased to 9.2 and incubated for additional 2 hours. Then reaction was neutralized and compound was purified by reverse-phase HPLC. Yield 89%. ESI-MS (m/z): $[M+H]^+$ calcd. for $C_{15}H_{28}N_6O_5$, 372.21; found, 372.2.

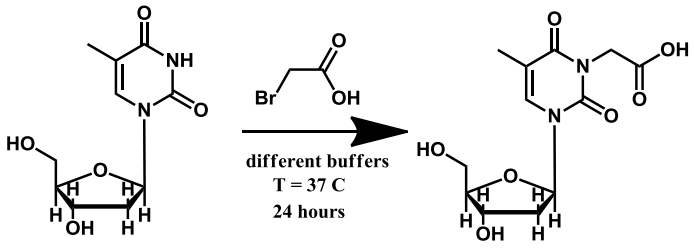
7.3 Model reactions with mononucleotides

7.3.1 Alkylation of 2-deoxyguanosine with 2-bromoacetate at different pH values.



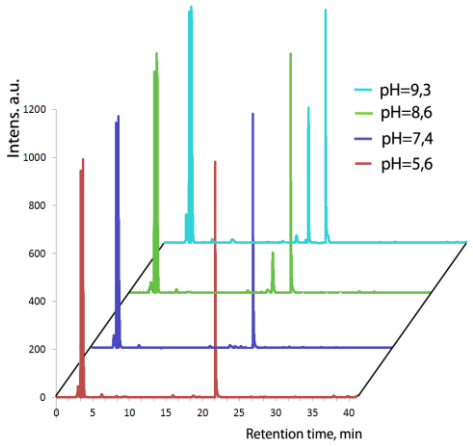
2.83 mg of guanosine (0.01 mmol) was dissolved in 100 µL solution of 1 M NaCl and 0.1 M of appropriate buffer (NH_4OAc, pH = 5.6; MOPS, pH = 7.4; borate buffer, pH = 8.6; carbonate buffer, pH = 9.3). In the independent flask 13.7 mg (10 eq., 0.1 mmol) of bromoacetic acid was dissolved in 100 µL water and pH was adjusted to buffer value with 100 mM NaOH. Then two solutions were mixed and incubated at 37 °C for 24 hours. Reverse-phase HPLC chromatogram was done with "HPLC method 2".

7.3.2 Modification 2-deoxythymidine with 2-bromoacetate at different pH values.



2.42 mg of 2-deoxythymidine (0.01 mmol) was dissolved in 100 µL solution of 1M NaCl and 0.1 M of appropriate buffer (NH_4OAc, pH = 5.6; MOPS, pH = 7.4; borate buffer, pH = 8.6; carbonate buffer, pH = 9.3). In the independent flask 13.7 mg (10 eq., 0.1 mmol) of bromoacetic acid was dissolved in 100 µL water and pH was adjusted to buffer value with 100 mM NaOH. Then two solutions were mixed and incubated at 37 °C for 24 hours. Reverse-phase HPLC chromatogram was done with "HPLC method 2". LC-MS measurements of the sample of incubation thymidine in borate buffer were performed in order to identify the compounds via their mass. HPLC chromatograms recorded at 260 nm of a samples incubated at pH = 5.6 and pH = 9.3 are depicted. Below the mass spectra of the peaks corresponding to nonmodified thymidine (rt = 21.2 min, 242.1 m/z) and to modified thymidine (rt = 16.8 min, $[M+Na]^+$ 323.0; $[M-H]^-$ 299.3 m/z) are shown. LC-MS conditions: flow rate: 0.3 mL/min; solvent A: H_2O + 0.1 %

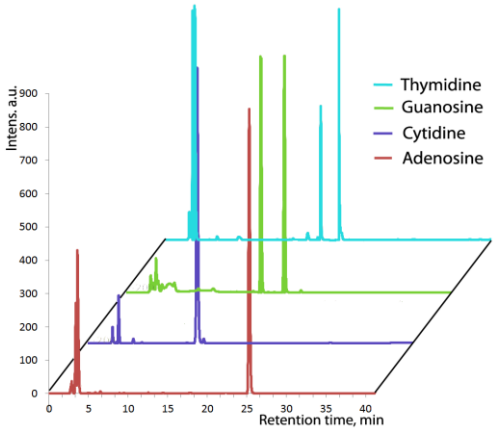
formic acid, solvent B: MeOH; gradient: 0-30 min 0-40% B, 30-39 min 40-100 % B. Yields in modification: pH = 5.6 – 0%, pH = 7.4 – 1.5%, pH = 8.6 – 12%, pH = 9.3 – 37%.



Supplementary figure S7.1. Reverse-phase HPLC chromatograms of alkylation of 2-deoxythymidine with 2-bromoacetic acid in different buffers.

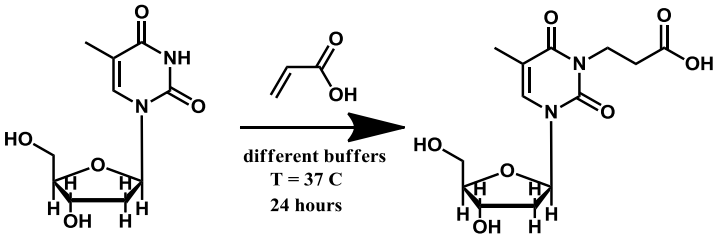
7.3.3 Modification of four nucleosides with 2-bromoacetate in sodium hydrogencarbonate buffer at pH 9.3.

0.01 mmol of mononucleotide (2-deoxyguanosine, 2.83 mg; 2-deoxythymidine, 2.42 mg; 2-deoxyadenosine, 2.51 mg; 2-deoxycytidine, 2.27 mg) was dissolved in 100 µL solution of 1M NaCl and 0.1 M Carbonate buffer, pH = 9.3. In the independent flask 13.7 mg (10 eq., 0.1 mmol) of bromoacetic acid was dissolved in 100 µL water and pH was adjusted to pH = 9.3 with 100 mM NaOH. Then two solutions were mixed and incubated at 37 °C for 24 hours. Reverse-phase HPLC chromatogram was done with "HPLC method 2". Yileds in modification: Adenosine – 0%, Cytidine – 0%, Guanosine – 53%, Thymidine – 41%.

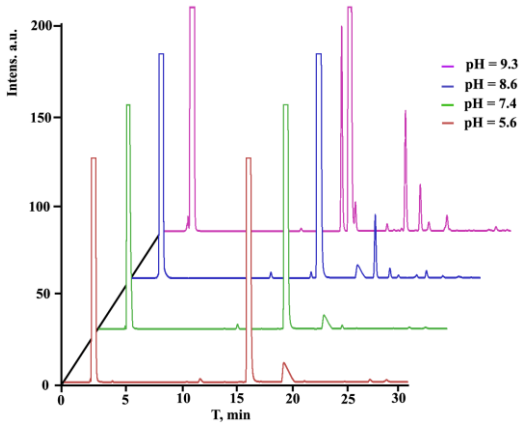


Supplementary figure S7.2. Reverse-phase HPLC chromatograms of alkylations of four deoxynucleotide under basic conditions.

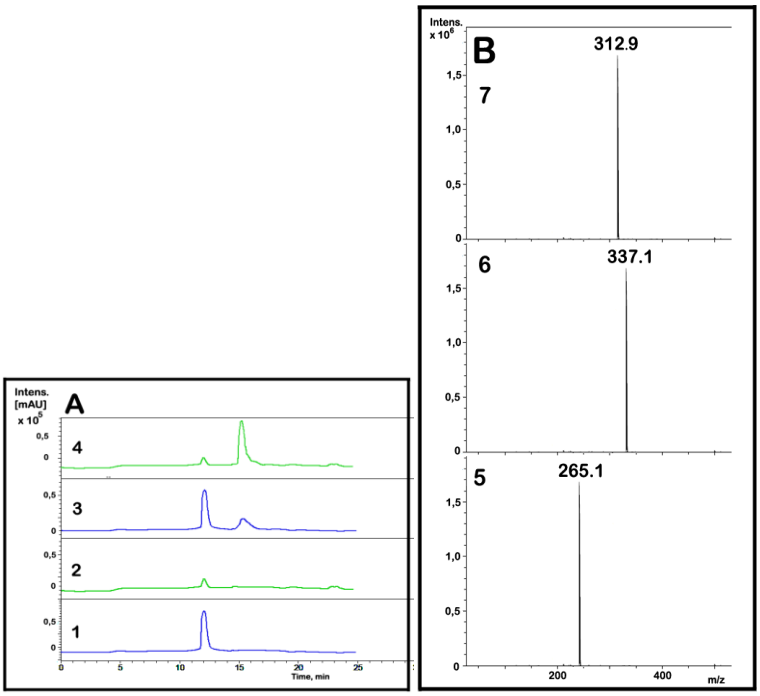
7.3.4 Modification of 2-deoxythymidine with acrylic acid at different pH values.



2.42 mg of thymidine (0.01 mmol) was dissolved in 100 µL solution of 1 M NaCl and 0.1 M of appropriate buffer (NH_4OAc, pH = 5.6; MOPS, pH = 7.4; sodium borate buffer, pH = 8.6; sodium carbonate buffer, pH = 9.3). In the independent flask 7.2 mg (10 eq., 0.1 mmol) of acrylic acid was dissolved in 100 µL water and pH was adjusted to buffer value with 100 mM NaOH. Then the two solutions were mixed and incubated at 37 °C for 24 hours. Reverse-phase HPLC chromatogram was done with "HPLC method 2". LC-MS measurements of the model reaction were performed in order to identify the compounds via their mass. HPLC chromatograms recorded at 260 nm of a samples incubated at pH = 5.6, pH = 7.4, pH = 8.6 and pH = 9.3 are depicted. Below the mass spectra of the peaks corresponding to nonmodified thymidine (rt = 12.5 min, 242.1 m/z) and to modified thymidine (rt = 15.8 min, $[M+Na]^+$ 337.2; $[M-H]^-$ 312.9 m/z) are shown. LC-MS measurements of the sample of incubation thymidine in borate buffer were performed in order to identify the compounds via their mass. Below the mass spectra of the peaks corresponding to nonmodified thymidine (rt = 16.8 min, $[M+Na+H]^+$ 265.1 m/z) and to modified thymidine (rt = 21.2 min, $[M*+Na]^+$ 337.1; $[M-H]^-$ 312.9 m/z) are shown. LC-MS conditions: flow rate: 0.3 mL/min; solvent A: H_2O + 0.1 % formic acid, solvent B: MeOH; gradient: 0-20 min 0-40% B, 20-39 min 40-100 % B. Yileds in modification: pH = 5.6 – 0%, pH = 7.4 – 0%, pH = 8.6 – 0.9%, pH = 9.3 – 19%.

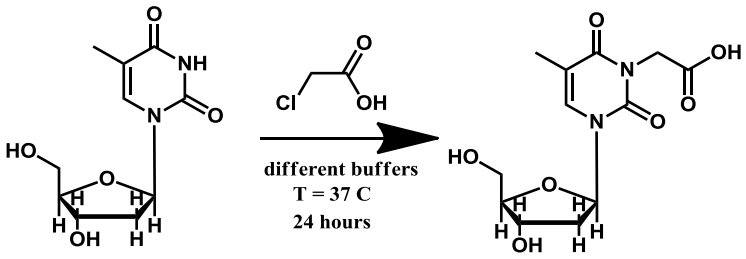


Supplementary figure S7.3. Reverse-phase HPLC chromatograms of alkylation of 2-deoxythymidine with acrylic acid in different buffers.



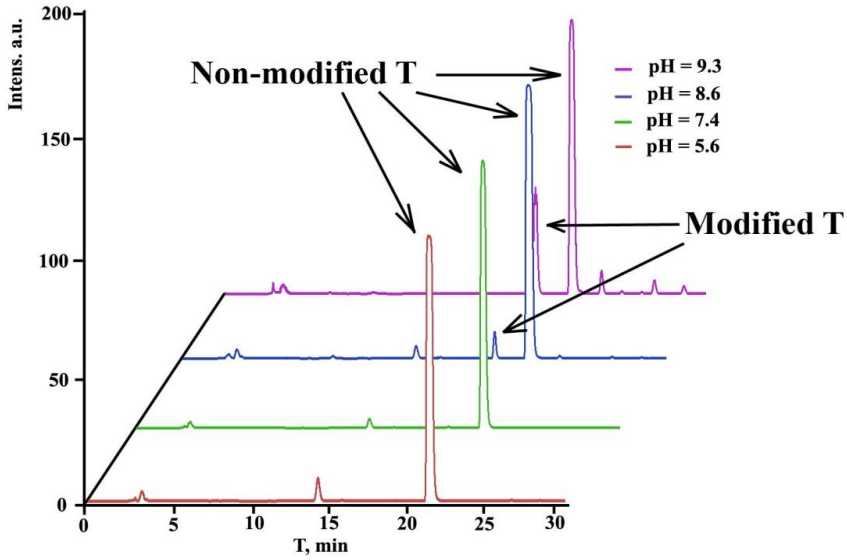
Supplementary figure S7.4. **A**: Reverse-phase HPLC chromatogram with different detectors (λ = 260nm – blue line, ESI negative mode – green line) of **1,2**: reaction of thymidine with bromoacetate at pH = 5.6; **3,4**: reaction of thymidine with bromoacetate at pH = 9.3. **B**: Deconvoluted ESI-MS spectra of **5**: Thymidine eluted at 12.5 min ($[M+H+Na]^+_{calc.}$ mass 265.2 Da), **6**: modified thymidine eluted at 15.8 min in positive mode ($[M+H+Na]^+$calc. 337.16 Da), **7**: modified thymidine eluted at 15.8 min in negative mode ($[M-H]^-$calc. 312.9 Da).

7.3.5 Modification of 2-deoxythymidine with 2-chloroacetic acid at different pH values.

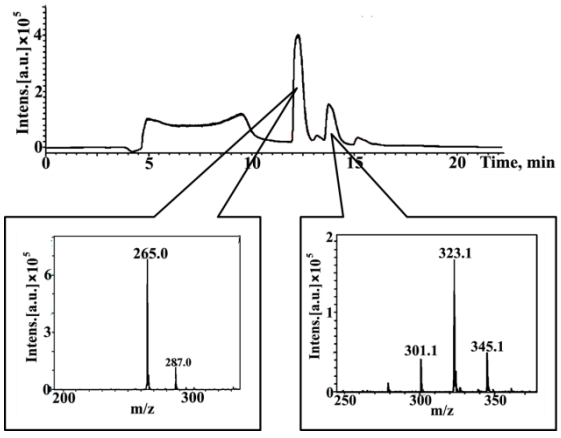


2.42 mg of 2-deoxythymidine (0.01 mmol) was dissolved in 100 µL solution of 1 M NaCl and 0.1 M of appropriate buffer (NH_4OAc, pH = 5.6; MOPS, pH = 7.4; sodium borate buffer, pH = 8.6; sodium carbonate buffer, pH = 9.3). In the independent flask 9.3 mg (10 eq., 0.1 mmol) of 2-chloroacetate dissolved in 100 µL water and pH was adjusted to buffer value with 100 mM

NaOH. Then two solutions were mixed and incubated at 37 °C for 24 hours. Reverse-phase HPLC chromatogram was done with "HPLC method 2". LC-MS measurements of the sample of incubation thymidine in borate buffer were performed in order to identify the compounds via their mass. HPLC chromatograms recorded at 260 nm of a samples incubated at pH = 5.6, pH = 7.4, pH = 8.6 and pH = 9.3 are depicted (**Fig. S7.5**). Below the mass spectra of the peaks corresponding to nonmodified thymidine (rt = 16.8 min, 242.1 m/z) and to modified thymidine (rt = 21.2 min, $[M+Na]^+$ 323.0; $[M-H]^-$ 299.3 m/z) are shown (**Fig. S7.6**). LC-MS conditions: flow rate: 0.3 mL/min; solvent A: H_2O + 0.1 % formic acid, solvent B: MeOH; gradient: 0-30 min 0-40% B, 30-39 min 40-100 % B. Yileds in modification: pH = 5.6 – 0%, pH = 7.4 – 0%, pH = 8.6 – 4%, pH = 9.3 – 25%.



Supplementary figure S7.5. Reverse-phase HPLC chromatograms of alkylation of 2-deoxythymidine with 2-chloroacetic acid in different buffers.



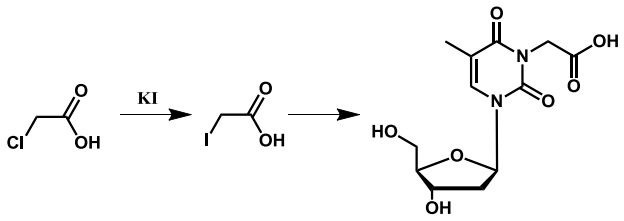
Supplementary figure S7.6: LC-MS analysis of N3 thymidine alkylation with 2-chloroacetate at 0.1 M $NaHCO_3$, pH = 9.3. Thymidine eluting at 13.5 min (calc. mass 265.09 Da), 2) modified thymidine eluting at 14.3 min in positive mode ($[M+H]^+_{calc.}$ = 300.16 Da)

($[M+H+Na]^+_{calc.}$ = 323.06 Da). Presence of additional peaks in both spectrum with molecular mass at 287.0 Da and 345.1 Da we correspond to attachment two sodium atoms to two carbonyl group in the heterocyclic core.

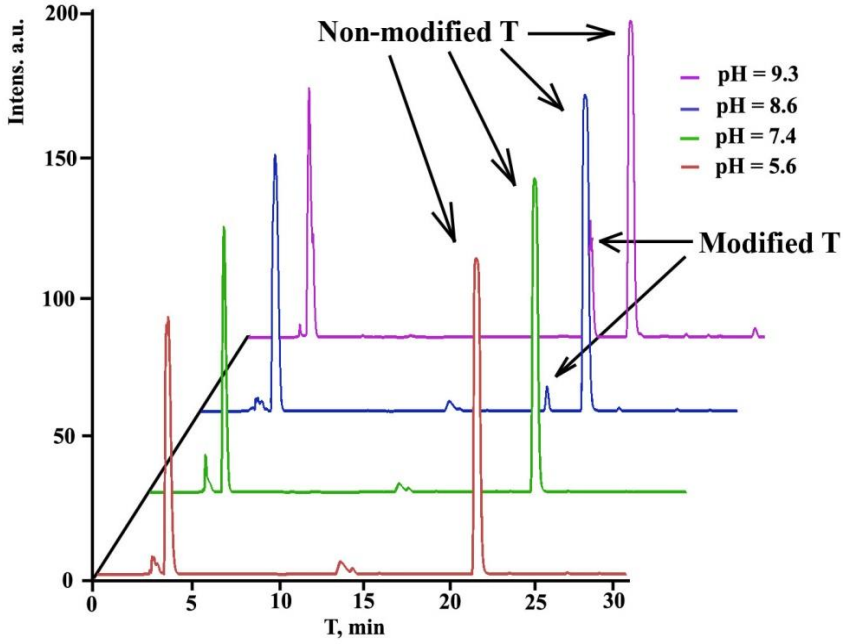
7.3.6 General protocol for mononucleotide alkylation with 2-chloroacetic acid.

0.01 mmol of mononucleotide (2-deoxyguanosine, 2-deoxythymidine, 2-deoxyadenosine or 2-deoxycytidine) was dissolved in 100 μL solution of 1 M NaCl and 0.1 M of appropriate buffer (NH_4OAc, pH = 5.6; MOPS, pH = 7.4; sodium borate buffer, pH = 8.6; sodium carbonate buffer, pH = 9.3). In the independent flask 9.3 mg (10 eq., 0.1 mmol) of 2-chloroacetic acid was dissolved in 100 μL water and pH was adjusted to buffer value with 100 mM NaOH. Then two solutions were mixed and incubated at 37 °C for 24 hours. Reverse-phase HPLC chromatogram was done with "HPLC method 2".

7.3.7 Modification of 2-deoxythymidine with 2-chloroacetic acid in different pH values catalysed by sodium iodide.

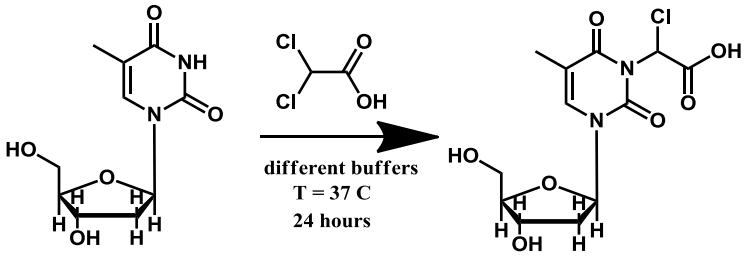


2.42 mg of thymidine (0.01 mmol) was dissolved in 100 μL solution containing 5 mM KI and 1 M NaCl in 0.1 M of appropriate buffer (NH_4OAc, pH = 5.6; MOPS, pH = 7.4; sodium borate buffer, pH = 8.6; sodium carbonate buffer, pH = 9.3). In the independent flask 13.7 mg (10 eq., 0.1 mmol) of bromoacetic acid was dissolved in 100 μL water and pH was adjusted to buffer value with 100 mM NaOH. Then two solutions were mixed and incubated at 37 °C for 24 hours. Reverse-phase HPLC chromatogram was done with "HPLC method 2". Yileds in modification: pH = 5.6 – 0%, pH = 7.4 – 0%, pH = 8.6 – 7%, pH = 9.3 – 31%.

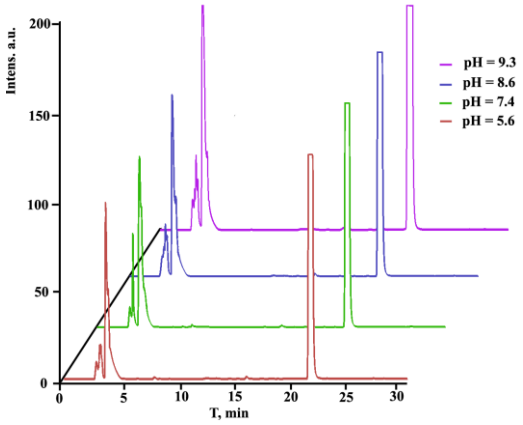


Supplementary figure S7.7. Reverse-phase HPLC chromatograms of alkylation of 2-deoxythymidine with 2-chloroacetic acid in different buffers in presence of NaI.

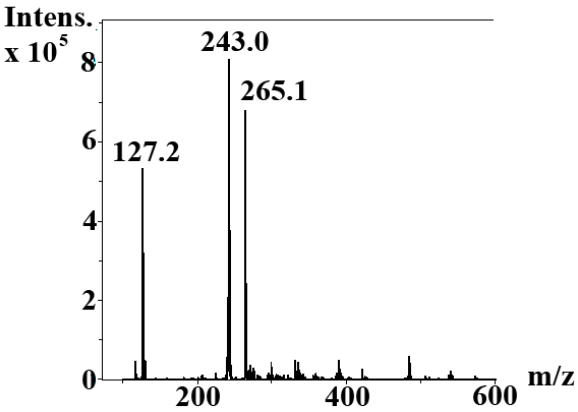
7.3.8 Modification of 2-deoxythymidine with 2,2'-dichloroacetic acid at different pH values.



2.42 mg of thymidine (0.01 mmol) was dissolved in 100 µL solution of 1 M NaCl and 0.1 M of appropriate buffer (NH_4OAc, pH = 5.6; MOPS, pH = 7.4; sodium borate buffer, pH = 8.6; sodium carbonate buffer, pH = 9.3). In the independent flask 12.7 mg (10 eq., 0.1 mmol) of 2,2'-dichloroacetic acid was dissolved in 100 µL water and pH was adjusted to buffer value with 100 mM NaOH. Then two solutions were mixed and incubated at 37 °C for 24 hours. Reverse-phase HPLC chromatogram was done with "HPLC method 2". LC-MS measurements of the sample of incubation thymidine in borate buffer were performed in order to identify the compounds via their mass. HPLC chromatogram recorded at 260 nm of a sample incubated at pH = 8.6 is depicted (**Fig. S7.8**). Below the mass spectra of the peaks corresponding to nonmodified thymidine (rt = 16.8 min, 242.1 m/z) is shown. LC-MS conditions: flow rate: 0.3 mL/min; solvent A: H_2O + 0.1 % formic acid, solvent B: MeOH; gradient: 0-30 min 0-40% B, 30-39 min 40-100 % B. Yileds in modification: pH = 5.6 – 0%, pH = 7.4 – 0%, pH = 8.6 – 0.1%, pH = 9.3 – 0.2%.

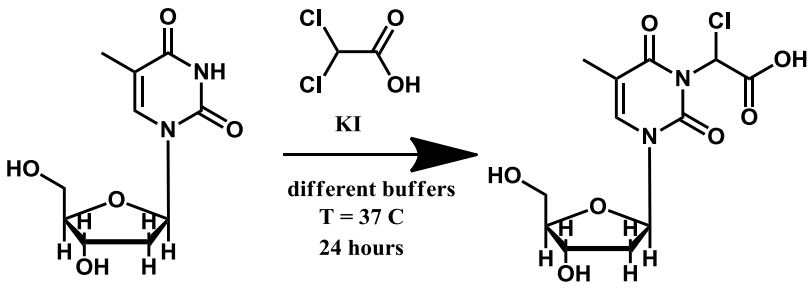


Supplementary figure S7.8. Reverse-phase HPLC chromatograms of alkylation of 2-deoxythymidine with 2,2'-chloroacetic acid in different buffers.

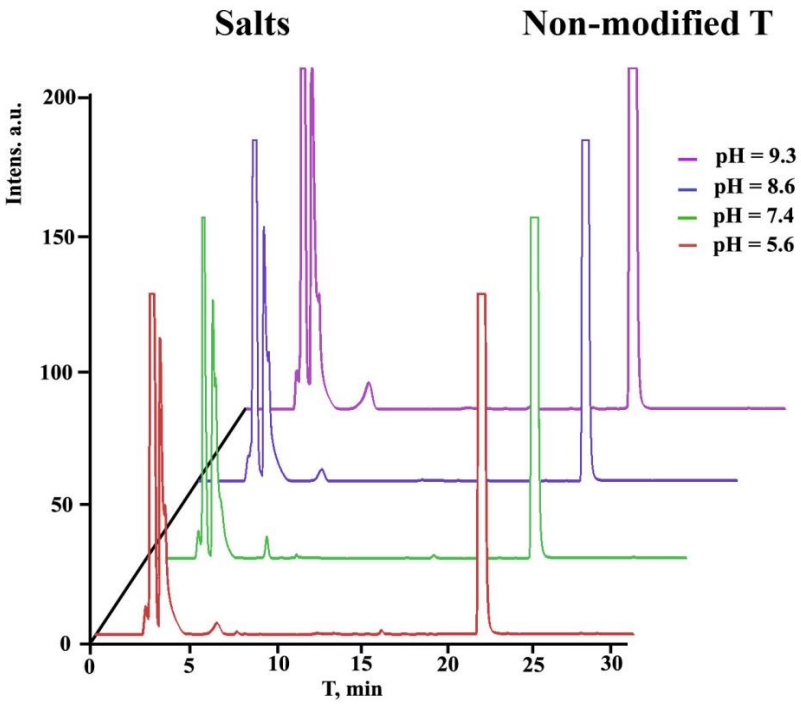


Supplementary figure S7.9. ESI-MS of reaction mixture of 2,2'-dichloroacetic acid pH = 9.3.

7.3.9 Modification of 2-deoxythymidine with 2,2'-dichloroacetic acid at different pH values catalysed by sodium iodide.

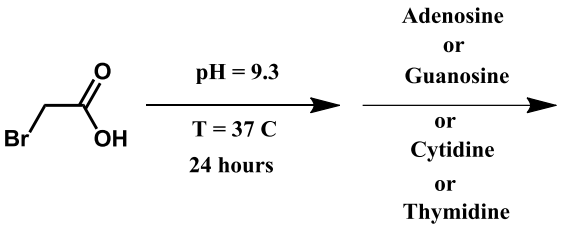


2.42 mg of thymidine (0.01 mmol) was dissolved in 100 µL solution containing 5 mM KI, 1 M NaCl and 0.1 M of appropriate buffer (NH_4OAc, pH = 5.6; MOPS, pH = 7.4; sodium borate buffer, pH = 8.6; sodium carbonate buffer, pH = 9.3). In the independent flask 12.7 mg (10 eq., 0.1 mmol) of 2,2'-dichloroacetic acid was dissolved in 100 µL water and pH was adjusted to buffer value with 100 mM NaOH. Then two solutions were mixed and incubated at 37 °C for 24 hours. Reverse-phase HPLC chromatogram was done with "HPLC method 2". Yields of modification in every reactions around 0%.



Supplementary figure S7.10. Reverse-phase HPLC chromatograms of alkylation of 2-deoxythymidine with 2,2'-chloroacetic acid in different buffers in presence of NaI.

7.3.10 Inactivation 2-bromoacetate in alkaline conditions.



13.7 mg (10 eq., 0.1 mmol) of bromoacetic acid was dissolved in 100 μL solution of 1 M NaCl and 1 M sodium carbonate buffer, pH = 9.3 at 37 °C for 24 hours. Then the appropriate mononucleotide (2-deoxyguanosine, 2.83 mg, 0.01 mmol; 2-deoxythymidine, 2.42 mg, 0.01 mmol; 2-deoxyadenosine, 2.51 mg, 0.01 mmol; 2-deoxycytidine, 2.27 mg, 0.01 mmol) was added to the solution of bromoacetate and incubated for additional 24 hours at 37 °C. Reverse-phase HPLC chromatogram was done with "HPLC method 2".

7.4 Modifications of DNA primers

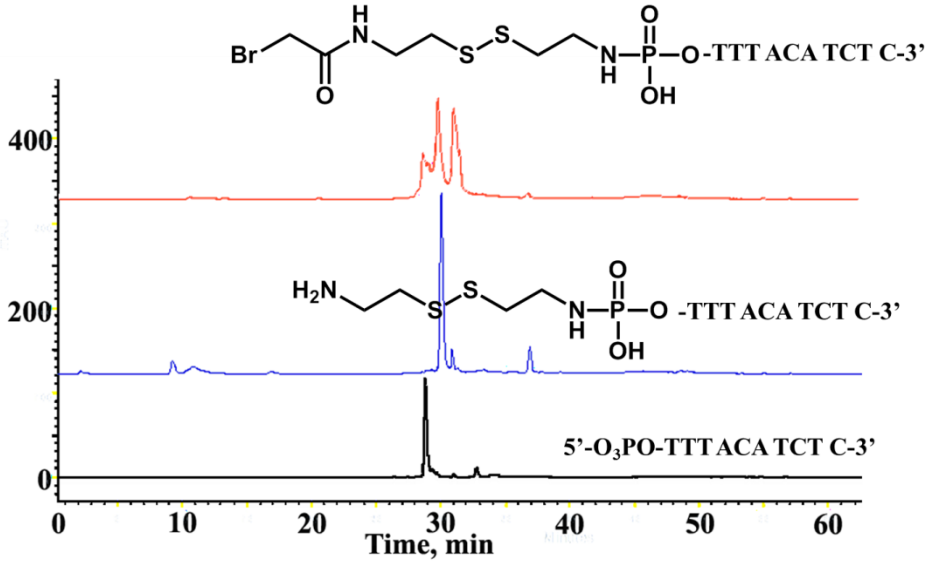
7.4.1 General procedure for attachment of amines or hydrazides to the 5'-phosphate group of oligonucleotides in organic and aqueous solutions.

The protocol of Boutorine et al[2] was followed, 100 nmol of cetyl trimethylammonium bromide (CTAB) salt of the oligonucleotide and 40 μmol of 4-dimethylaminopyridine were dissolved in 50 μL of DMSO and mixed with 30 μmol of dipyridyl disulfide in 25 μL of DMSO and 30 μmol of triphenylphosphine in 25 μL of DMSO. After 30 min. of incubation at 37 °C a solution of cystamine (40 μmol) in DMSO was added and incubated at the same temperature for the next 2 hours and the conjugate was twice precipitated with 3 volumes of EtOH, and washed with 70% EtOH. Analysis of reaction by HPLC and MALDI-TOF MS show complete conversion initial oligonucleotide to the disulfide containing molecule without significant amounts of side products. Reactive sequences that were synthesized according to this protocol: SRS_3.12, SRS_3.14, SRS_3.15, SRS_3.21, SRS_3.22, SRS_3.23, SRS_3.24, SRS_3.29, SRS_3.30, SRS_5.12, SRS_5.13.

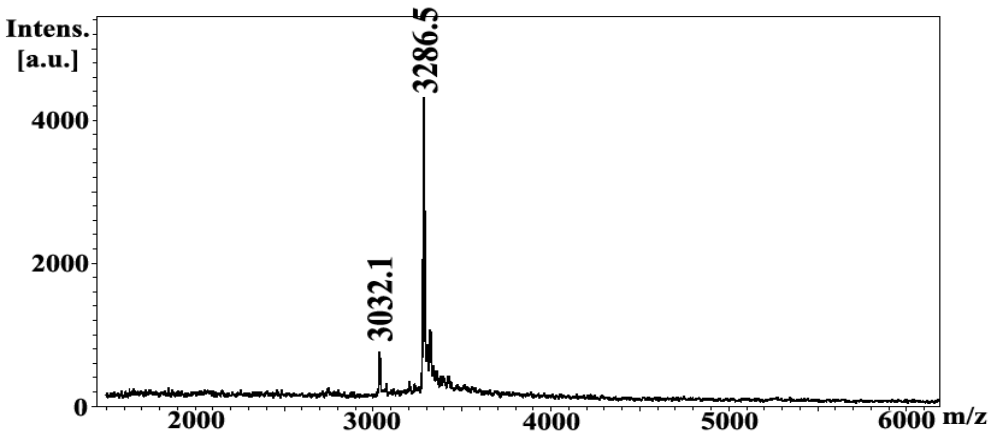
7.4.2 General procedure for acetylation of amino-containing DNA with N-hydroxysuccinimido-2-bromoacetate.

Amino-containing oligodeoxyribonucleotide sequence (100 μmol in 200 μL of 200 mM sodium tetraborate buffer, pH 8.7) was treated with 1 mM *N*-hydroxysuccinimidyl bromoacetate in DMSO (200 μL) at 37 °C for 30 min. The (*N*-bromoacetyl)oligonucleotide product was precipitated with 0.3 M sodium acetate, pH 5.2, in 3 volumes of EtOH, washed with 70% EtOH, and then dissolved in water. The solution was desalted using NAP-5 size exclusion column

(Amersham Biosciences). The (*N*-bromoacetyl)-oligonucleotide (eluted from the column with 1 mL of deionized, distilled water) was quantified by UV/visible spectrophotometry and immediately used in the next step. Analysis was done by analytical reverse-phase HPLC using "HPLC method 3". Reactive sequences that were synthesized according to this protocol: SRS_4.2, SRS_4.4, SRS_4.6, SRS_4.8, SRS_4.10, SRS_4.12, SRS_4.13, SRS_5.2, SRS_5.4, SRS_5.5, SRS_5.6, SRS_5.8, SRS_5.9, SRS_5.11, SRS_5.15, SRS_6.1, SRS_2.2.



Supplementary figure S7.11. Example of reverse-phase HPLC chromatograms of synthesis of reactive sequence (SRS_4.2).

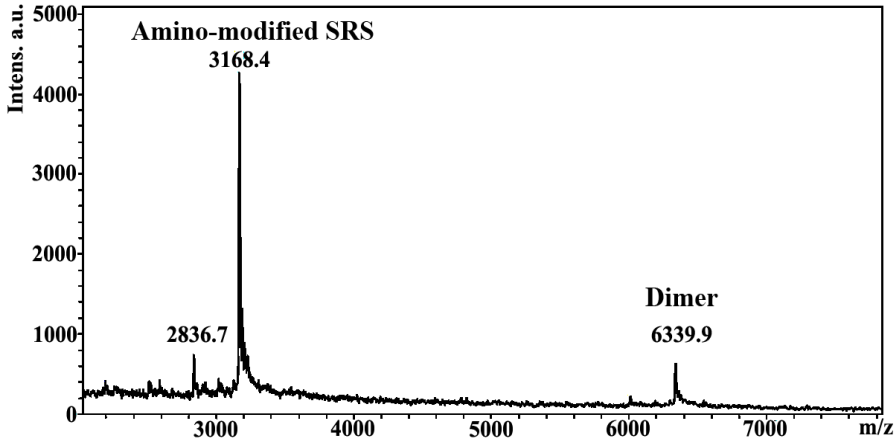


Supplementary figure S7.12. Example of MALDI-TOF MS of reactive sequence (SRS_4.2) $M_{calc.}$ = 3284.9.

7.4.3 General procedure for the transamination of cytidine in the recognition sequence by amines or hydrazides.

0.1 mL of concentrated HCl, 0.1 mL of water and 0.1 mL of ethylenediamine were mixed and the pH was adjusted to pH = 7.5. Then 0.186 g of sodium hydrogensulfite was added and pH of solution was decreased to 5.6 with 0.01 M HCl. The final volumes were brought to 1 mL with distilled water. A 1 mg/mL portion of hydroquinone (dissolved in 95% ethanol) was added to the sodium hydrogensulfite solutions to scavenge free radicals. The reaction was initiated by adding 0.1 mL of 100 μM DNA SRS to the transaminated mixture. The reaction mixture was incubated for 24 h at 37°C. Then pH was adjusted to 9.6 and incubated for additional 2 hours at 37 °C. After an overnight dialysis against 5 mM sodium phosphate buffer, pH 7.5, the samples were lyophilized. Yield of conversion is quantative (in case of ethylenediamine M_{calc} = 3169.2, M_{found} = 3168.4). Similar protocol was applied for other type of diamine linkers.

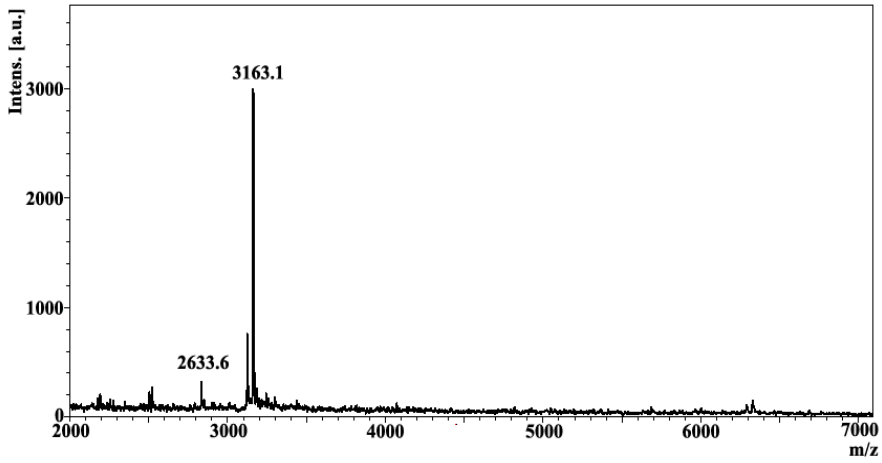
Reactive sequences that were synthesized according to this protocol: SRS_3.3, SRS_3.4, SRS_3.5, SRS_3.6, SRS_3.7, SRS_3.8, SRS_3.9, SRS_3.16, SRS_3.17, SRS_3.18, SRS_3.19, SRS_3.23, SRS_3.24, SRS_3.25, SRS_3.26, SRS_3.27, SRS_3.28, SRS_3.32, SRS_3.33, SRS_3.34, SRS_3.35, SRS_3.36, SRS_3.37,



Supplementary figure S7.13. Example of MALDI-TOF MS of reactive sequence (SRS_3.3) M_{calc} = 3169.2, M_{found} = 3168.4.

7.4.4 General procedure for the transamination of cytidine in the recognition sequence by propylamine

Transamination with propylamine was done same to protocol 7.4.3. Yield of conversation is 65%.

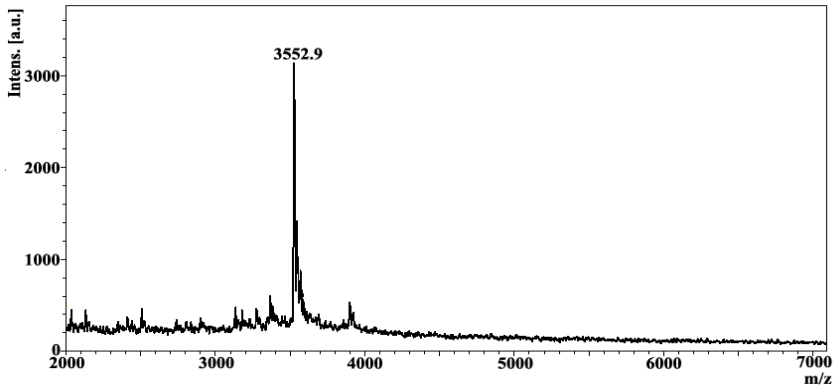


Supplementary figure S7.14. Example of MALDI-TOF MS of reactive sequence (SRS_3.5) M_{calc} = 3163.8, M_{found} = 3163.1.

7.4.5 Coupling of diamine linker to carboxylic-containing recognition sequence.

2.5 pmol of ethylenediamine were dissolved in 20 μL of H_2O. and 5 μL of 1.2 mM water solution of carboxylic-linked oligonucleotide (SRS_3.38) (6 nmoles) were added. The solution was cooled to 0°C and 1.3 mg (6.8 pmol) of EDC in 1 μL of water were added. The mixture was incubated at 37°C for 2 hours. Then one more portion of EDC was added and mixture was incubated for additional 2 hours. The oligonucleotide was isolated by precipitation with ethanol/sodium acetate and purified by RP-HPLC using "HPLC method 3". Yield of conversion is 54%. MALDI-TOF MS M_{calc} = 3556.0, M_{found} = 3552.9

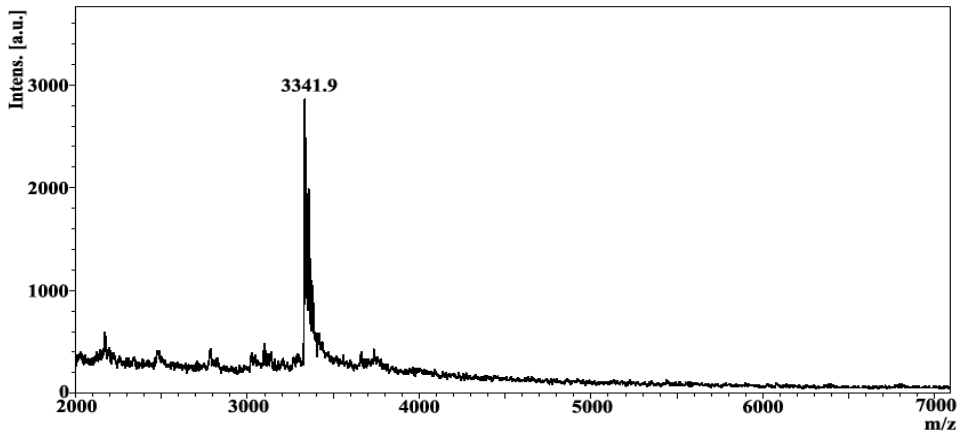
Reactive sequences that were synthesized according to this protocol: SRS_3.11, SRS_3.13, SRS_3.20, SRS_3.23.



Supplementary figure S7.15. MALDI-TOF MS of reactive sequence SRS_3.39.

7.4.6 Deprotection of benzoyl-protecting group (SRS_3.6).

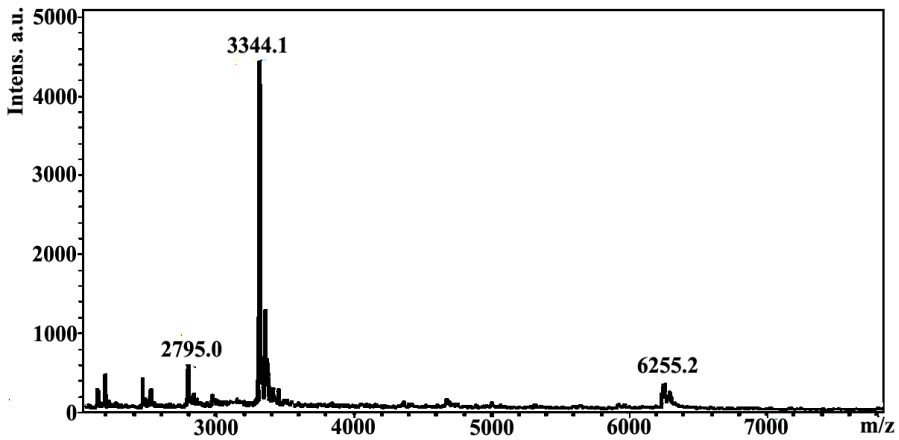
Benzoil-containing DNA sequence SRS_3.39 was added to 0.5 M phosphate buffer pH = 8.6 that contains 10 mM of piperidine. Reaction was incubated for 6 hours at 37 °C and precipitated by ethanol. Further purification by RP-HPLC using "HPLC method 5" produced deprotected SRS_3.6 in 68% yield. MALDI-TOF MS M_{calc} = 3344.1, M_{found} = 3341.9



Supplementary figure S7.16. MALDI-TOF MS of reactive sequence SRS_3.6.

7.4.7 Introducing a diol-cleavage site by transamination of cytosine with a tartaric-derivative (SRS_3.6).

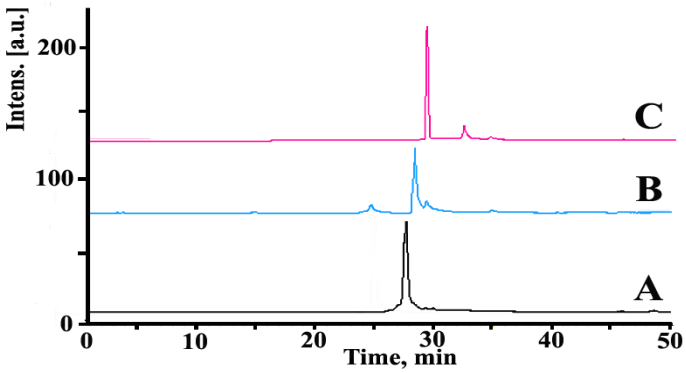
0.1 g of N,N'-Di-(2-aminoethyl)-tartaramide (0.4 mmol) was added to a solution containing 0.186 g of sodium hydrogensulfite and pH of solution was adjusted to 5.6 with 0.01 M HCl. The final volumes were brought to 1 mL with distilled water. A 1 mg/mL portion of hydroquinone (dissolved in 95% ethanol) was added to the sodium hydrogensulfite solutions to scavenge free radicals. The reaction was initiated by adding 0.1 mL of 100 μM DNA SRS_3.2 to transaminated mixture. The reaction mixture was incubated for 24 h at 37°C. Then pH was adjusted to 9.6 and incubated for additional 2 hours at 37 °C. After an overnight dialysis against 5 mM sodium phosphate buffer, pH 7.5, the samples were purified by RP-HPLC using "HPLC method 3". MALDI-TOF MS M_{calc} = 3344.1, M_{found} = 3341.9. Yield of conversion is 60%.



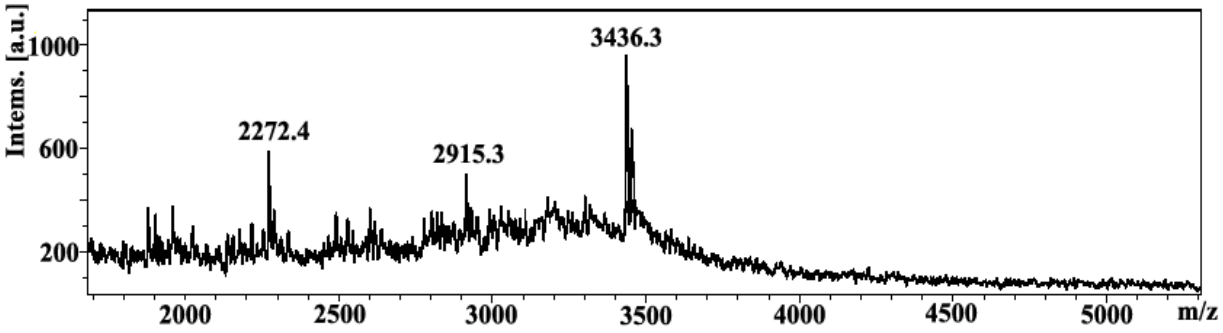
Supplementary figure S7.17. MALDI-TOF MS of reactive sequence SRS_3.6.

7.4.8 Introducing a diol-cleavage site *via* "click" chemistry.

Alkyno-modified DNA SRS_3.5 (10 nmol) was dissolved in 100 µL of 0.5 M Tris buffer pH = 7.4. To this solution was added sodium ascorbate (500 nmol) in 50 µL of water, 500 nmol of N-(2-aminoethyl)-N'-(propan-3-azide)-O,O'-L-tartaramide in 50 µL of water, and 10 nmol of tris[(1-benzyl-1H-1,2,3-triazol-4-yl)methyl]amine in 100 µL of DMSO. Reaction was activated by adding 1 nmol of $CuSO_4$ in 10 µL of water. Reaction was incubated under nitrogen at 25 °C for 12 hours. Then DNA were precipitated with 0.3 M sodium acetate, pH 5.2, in 2.5 volumes of EtOH, and washed with 70% EtOH. Analysis by analytical reverse-phase HPLC using a gradient of 1-40% CH_3CN (1.00%/min) in 0.05 M TEAA, pH 7.0, indicated clean conversion of starting amine to product (85%). MALDI-TOF MS M_{calc} = 3436.7, M_{found} = 3436.3. Reactive sequences that were synthesized according this protocol: SRS_3.10, SRS_3.34.



Supplementary figure S7.18. Example of reverse-phase HPLC chromatograms of synthesis of reactive sequence (SRS_3.10). **A**: initial primer; **B**: alkyne-modified SRS_3.5; **C**: SRS_3.10.



Supplementary figure S7.19. MALDI-TOF MS of reactive sequence SRS_3.15.

7.4.9 PNA synthesis (synthesis was done according published procedure[3] with some modifications).

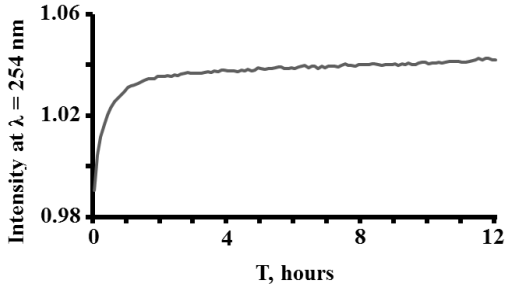
PNAs were synthesized by solid-phase peptide synthesis using Fmoc/Bhoc-protected PNA monomers (Applied Biosystems) and Fmoc-protected PALPEG-PS resin (Peptides International, Louisville, KY) with a 0.25 mmol/g loading capacity. 100 mg of resin was shaken gently in dichloromethane (DCM) for several hours to expose reactive sites on the resin prior to peptide

coupling. The resin was then deprotected three times with 20% piperidine in *N*,*N*-dimethylformamide (DMF) for 5 min to remove the Fmoc protecting group. A 5-fold excess of PNA monomer (based on manufacturer's reported loading capacity) in basic solution (Applied Biosystems) and activator solution (Applied Biosystems) was then prepared. The base solution consists of a mixture of 2,6-lutidineand*N*,*N*-diisopropylethylamine in DMF. The activator is a solution of *O*-(7-azabenzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HATU) in DMF. Following a 10-min activation step, the solution was added to the resin. The initial monomer coupling to resin was allowed to proceed for at least 2 h to ensure complete coupling on the optimal number of resin sites. The resin was washed with DCM and DMF following the coupling step and a small fraction collected for a cianine test. Unreacted sites were then coupled with additional amount of appropriate monomer until negative reaction on free aminogroup were observed. The resin was washed with DCM and DMF, and the whole process of deprotection and coupling was repeated until the desired PNA peptide was synthesized. The resin was soaked in a mixture of 4:1 trifluoroacetic acid (TFA)/TIS/H_2O 97/2/1 (total volume 1 mL) for 2 h to cleave the PNAA from the resin and remove Bhoc side protecting groups. Once cleaved from the support, the PNA was precipitated by the addition of dry ether (10- fold excess over PNA). The solid was separated from the ether by centrifugation (1000*g*) for 5 min. Upon completion, the PNA pellet was dried within a stream of nitrogen and dissolved in deionized water. Semipreparative-scale PNA purification was performed on a Waters Delta 600 HPLC using a Zorbax RP-C4 column with a particle size of 5 *i*m. Elution of the PNAA product was achieved using a 40-min linear gradient (0% - 80% Acetonitrile, 0.1% TFA in water) and a 4 mL/min flow rate. The yield of the reaction was 30% (based on the manufacturer's reported resin loading), or 0.0075 mmol of PNA product. Structure of desir PNA was proved by MALDI-TOF. Reactive sequences that were synthesized according this protocol: PNA_4.1, PNA_4.2, PNA_4.3.

7.4.10 Evaluation of the influence of the sulfone group on the stability of the DNA duplex during transamination of cytidine.

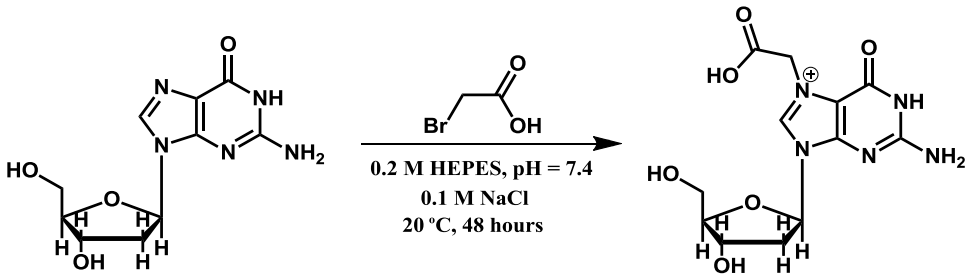
Template single DNA strand with 5 cytidines (ODN_3.4) was incubated for 6 hours with 1 M $NaHSO_3$ at pH = 5.6 and then dialyzed against 5 mM sodium acetate pH = 5.6. Sample was lyophilized and a 100 μM stock solution in 1 M NaCl, 0.1 mM sodium acetate buffer at pH = 5.6was prepared. Sulfonated template and complementary SRS_3.2 was mixed in equimolar ratio and UV spectrophotometric determination of melting curves was done with heating speed 1°C. After finishing, pH of solution were adjusted to pH = 9.3 with 1 M sodium carbonate buffer and increasing intensity at 254 nm was monitored (**Fig. S7.20**). After 12 hours pH of solution

was adjusted to 7.4 and hybridization features of the duplex were again determined with UV spectrophotometric melting curves.



Supplementary figure S7.20. Kinetic of changing intensity band at 254 nm during hydrolysis of sulfone group from cytidine core.

7.4.11 Evaluation of stability of DNA strand under model conditions.



7.4.11.1 α-Bromoacetic acid.

α-Bromoacetic acid ($13^{\times}10^{-3}$g, 100μmol, 50 eq.) was dissolved in 50 μL of 0.4 M HEPES pH = 7.4, 0.2 M NaCl and pH was adjusted to 7.4. Then 20 μL of 100 μM DNA stock solution was added. Mixture was adjusted by water to 100 μL. Reaction was incubated at 37°C. Aliquots were taken out after 0 h, 2 h, 6 h, 12 h, 24 h. HPLC traces: 24 min – 100% (initial peak). The solvents consisted of 100 mM ammonium acetate buffer adjusted to pH 5.6 using acetic acid (solvent A) and pure methanol (solvent B). The elution started with isocratic flow of 100% solvent A for 3 min, followed by a linear gradient to 30% solvent B at 10 min, then to 70% solvent B at 20 min and further to 80% solvent B at 25 min. Initial conditions were regenerated by rinsing with 100% solvent A for 10 min. The flow rate was 0.3 mL/min. MALDI: 6116.8 Da (initial), 6174.8 (+59 Da, modified G) (see chapter 4.2).

7.4.11.2 TCEP

TCEP solution (100 μmol, 50 eq.) was added to 50 μL of 0.4 M HEPES pH = 7.4, 0.2 M NaCl and pH was adjusted to 7.4. Then 20 μL of 100 μL DNA stock solution was added. Mixture was adjusted by water to 100 μL. Reaction was incubated at 37°C. Aliquots were taken out after 0 h, 2 h, 6 h, 12 h, 24 h and analysed as above. No degradation of DNA was observed.

7.4.11.3 α-Bromoacetic acid and TCEP.

α-Bromoacetic acid ($13^{\times}10^{-3}$g, 100 μmol, 50 eq.) was dissolved in 50 μL of 0.4 M HEPES pH = 7.4, 0.2 M NaCl and pH was adjusted to 7.4. Then 20 μL of 100 μL DNA stock solution was added. Mixture was adjusted by water to 100 μL. Reaction was incubated at 37 °C for 12 hours. Then 20 μL of TCEP (200 mmol) were added and incubation was continued. Aliquots were taken out after 0 h, 2 h, 6 h, 12 h, 24 h. HPLC traces: 24 min – 100% (initial peak). MALDI: 6116.8Da (initial), 6174.8 (+59 Da, adduct) (see chapter 4.2).

7.4.11.4 α-Bromoacetic acid with piperidine cleavage.

α-Bromoacetic acid ($13^{\times}10^{-3}$g, 100 μmol, 50 eq.) was dissolved in 50 μL of 0.4 M HEPES pH = 7.4, 0.2 M NaCl and pH was adjusted to 7.4. Then 20 μL of 100 μM DNA stock solution was added. Mixture was adjusted by water to 100 μL. Reaction was incubated at 37°C for 24 hours. Then 5 μL of 1 M piperidine in H_2O was added to solution and heated to 90 °C. After 20 min, sample was cooled in ice water and evaporated under vacuum overnight.

7.5 Cross-linking protocols

To accelerate duplex formation between target and reactive sequences prior to the cross-linking reaction, the two strands in ration 2:1 reactive sequence and template, were mixed in buffer containing 1 M NaCl, heated for 1 minute at 50 °C and slowly cooled to room temperature.

7.5.1 Transamination protocol №1.

Reactive and template sequences were mixed in 1 M solution of sodium hydrogensulfite pH = 5.6 and incubated at RT for 7 days. Then pH of the solution was increased to 9.3, and the sample was incubated for two more hours at RT. Reaction was dialyzed overnight (or by Nap-5 size-exclusion column) and samples were analysed by PAGE. In case of radiolabeled samples, pH of reaction was adjucted to 7.4 and the sample purified by Zip-Tip.

7.5.2 Transamination protocol №2.

Template was preactivated with 1 M sodium hydrogensulfite for 12 hours. Then reactive sequence was added. Reaction was incubated for next 12 hours at RT. To stop reaction, pH of solution was increased to 9.3, and sample was incubated for two more hours at RT. Reaction was

dialyzed overnight (or by Nap-5 size-exclusion column) and samples were analysed by PAGE. In case of radiolabeled samples, pH of reaction was adjusted to 7.4 and the sample purified by Zip-Tip.

7.5.3 Protocol for alkylation of guanine.

1 μM reactive (*N*-bromoacetyl)oligonucleotide, 0.1 M MOPS, pH 7.4 in 1 μL, and 1 μM DNA-template were adjusted to a total volume of 10 μL. Reactions were incubated at 20 °C for 24 h. After Zip-Tip purification, cross-linked products were separated on 18% denaturing polyacrylamide gels.

7.5.4 Protocol for alkylation of thymidine.

1 μM reactive (N-bromoacetyl)oligonucleotide, 0.1 M sodium borate buffer, pH 8.6 in 1 μL, and 1 μM DNA-template were adjusted to a total volume of 10 μL. Reactions were incubated at 20 °C for 24 h. After Zip-Tip purification (or by Nap-5 size-exclusion column), cross-linked products were separated on 18% denaturing polyacrylamide gels.

7.5.5 Cross-link formation in triplex formation system.

10 nmoles of single stranded and complementary oligonucleotide templates were hybridized as follows: Oligonucleotides were combined in 40 μL of 0.1 M HEPES, 1 M NaCl, 5 mM $MgSO_4$. The solution was heated to 90 °C for 5 min and then cooled to room temperature. After 2 h, the resulting solution was then combined with 10 nmoles of SRS N-bromoacetyloligonucleotide . This solution was stirred at 4 °C or 20 °C for 24 h. After this aliquot of sample were desalted and loaded on 18% denaturing polyacrylamide gels.

7.6 DNA cleavages protocols

7.6.1 Piperidine cleavage was followed by previously published procedure[4].

A DNA sample containing N7-alkylated guanine was suspended in 100 μL of an ice-cold solution of 10% piperidine in water. Samples were heated at 90 °C for 15 minutes to convert quantitatively the sites of guanine-N7 alkylation into single strand breaks. Samples were then lyophilized to remove traces of piperidine.

7.6.2 Disulfide cleavage.

Target (10 nmol) and complementary (20 nmol) reactive single-stranded oligonucleotides were mixed. Then 5 μL of 200 mM stock solution of TCEP (pH = 7.2) was added and reaction was incubated for 2 hours. Zip-Tip purified and loaded directly to 18% denaturing polyacrylamide gels.

7.6.3 mBBr labelling

Target (10 nmol) and complementary (20 nmol) reactive single-stranded oligonucleotides were mixed (in 10 μL of 0.2 M MOPS pH = 7.4, 0.5 M NaCl) and incubated at 20 °C for 24 h. Then 5 μL of 200 mM stock solution of TCEP (pH = 7.2) was added and reaction was incubated under nitrogen in glove box at RT for 2 hours. Then 0.44 μL of 100 mM stock solution of mBBr was added and mixture was incubated at 25 °C for 30 min. Then reaction was desalting using Nap-5 column. Final solutions were analysed by RP HPLC "*DNA_Cy3_FLD*" (gradient of 5-10% CH_3CN (0.5 mL/min) in 0.05 M TEAA, pH 7.0) in 10 min. Then CH_3CN from 10% to 20% from 10 to 30 min, then from 30% to 80% from 30 min to 45 min. UV lamp with wavelength from 240 nm to 800 nm was used as a detector. Fluorescent detector with excitation wavelength at 390 nm and emission at 490 nm was used as additional detector.

7.6.4 Cleavage and labeling of diol-cleavage site.

Target template was incubated in 200 μL of 1 M $NaHSO_3$ (pH 5.6) for 12 hours at RT and 2 eq. of reactive sequence were added. After incubation for 12 hours, pH was increased up to 9.3 and reaction was incubated for 2 hours. Temperatures during the whole reaction were kept constant in an Eppendorf thermomixer comfort at 20 °C. Then reaction was desalted using NAP-5 column and lyophilized. The mixture of lyophilized DNA's was dissolved in 200 μL 0.5 M sodium carbonate at pH = 4.5. To this solution 10 μL of 1 M $NaIO_4$ was added and incubated at 37 °C for 2 hours. Unreacted sodium periodate was quenched by ethyleneglycol and amine-derivative (tris(hydroxymethyl)amino-methane or 5-(aminoacetamido)fluoresceine) was added. After 2 hours 100 μL of 0.1 M $NaCNBH_3$ was added. Reaction was left for one additional hour, precipitated by EtOH and analysed by RP-HPLC.

7.6.5 Cleavage of disulfide bridge and introducing cyanine dyes in the presence of TCEP.

Target (10 nmol) and complementary (20 nmol) reactive single-stranded oligonucleotides were mixed (in 10 μL of 0.2 M MOPS pH = 7.4, 0.5 M NaCl) and incubated at 20 °C for 24 h. Then 5 μL of 200 mM stock solution of TCEP (pH = 7.2) was added and reaction was incubated under nitrogen in glove box at RT for 2 hours. Then 0.44 μL of 100 mM stock solution of cyanine dye

(maleimido-sulfo-Cy5 or maleimido-sulfo-Cy3) was added and mixture was incubated at 25 °C for 30 min. Then reaction was desalted using Nap-5 column to remove any organic compounds. Final solutions were analysed by RP HPLC "*DNA_Cy3_FLD*" (gradient of 5-10% CH_3CN (0.5 mL/min) in 0.05 M TEAA, pH 7.0) in 10 min. Then CH_3CN from 10% to 20% from 10 to 30 min, then from 30% to 80% from 30min to 45 min. UV lamp with wavelength from 240 nm to 800 nm was used as a detector. Fluorescent detector with excitation wavelength at 390 nm and emission at 490 nm was used as additional detector.

7.6.6 Cleavage of disulfide bridge and introducing cyanine dyes in the abscence of TCEP.

Target (10 nmol) and complementary (20 nmol) reactive single-stranded oligonucleotides were mixed (in 10 μL of 0.2 M MOPS pH = 7.4, 0.5 M NaCl) and incubated at 20 °C for 24 h. Then 5 μL of 200 mM stock solution of TCEP (pH = 7.2) was added and reaction was incubated under nitrogen in glove box at RT for 2 hours. After this, sample was desalted by size-exclusion Nap-5 column in the glove box under N_2. Then 0.44 μL of 100 mM stock solution of cyanine dye maleimido-sulfo-Cy3 was added and mixture was incubated at 25 °C for 2 hours. Then reaction was desalted using Nap-5 column to remove any organic compounds. Final solutions were analysed by RP HPLC "*DNA_Cy3_FLD*" (gradient of 5-10% CH_3CN (0.5 mL/min) in 0.05 M TEAA, pH 7.0) in 10 min. Then CH_3CN from 10% to 20% from 10 to 30 min, then from 30% to 80% from 30min to 45 min. UV lamp with wavelength from 240 nm to 800 nm was used as a detector. Fluorescent detector with excitation wavelength at 390 nm and emission at 490 nm was used as additional detector.

7.7 Stability of cyanine dye

5 μL of 100 mM stock solution of cyanine dye (maleimido-sulfo-Cy5 or maleimido-sulfo-Cy3) and 5 μL of 0.2 M TCEP were added in 45 μL of 0.2 M MOPS pH = 7.2, 0.5 M NaCl. Reaction was incubated at RT for 2 hours and reaction mixture was directly injected to RP=HPLC.

7.8 Stability of N7-guanine-alkylated DNA strand.

7.8.1 Stability of DNA template with non-specifically modified guanine residues.

1 mg of 2-bromoacetic acid was dissolved in 50 μL water and pH was adjusted to 7.4 with 100 mM NaOH. To this solution was added 20 μL of 1 M MOPS buffer pH=7.4 and 30 μL of 5 M NaCl and 100 μL of 100 mM DNA stock solution. Reaction was incubated for 24 hours at 37 °C. Then solution was desalted by NAP-5 column to remove bromoacetic acid and left at RT. Aliquots were taken after 1 day, 2 days and 4 days. After 4 days piperidine cleavage was done.

All samples were analysed by “HPLC method 1”. Solvent “A”: TEAA 100 mM, pH = 7.0. Solvent “B”: Acetonitrile. Speed: 0.5 mL/min. Solvent gradient: from 0min to 20 min “B” from 0% to 5%; from 20min to 40 min – solvent “B” from 5% to 25%, from 40min to 50 min – from 25% to 80%.

7.8.2 PAGE analysis of cross-link stability.

DNA template was labelled with P^{32} and mixed with 2 eq. of SRS with reactive and cleaved module. Reaction was incubated for 24 hours in 0.1 M MOPS buffer, pH = 7.4, 1 M NaCl at RT. Solution was divided into three parts and reactive group was quenched in different ways. After quenching, samples were left at RT and aliquot was taken out after 1 day, 2 days and 4 days. As a control for formation of intermediate with cleavable bridge, each of three parts was treated with TCEP. First part was incubated in the initial buffer at pH = 7.4. In the second part pH of solution was decreased to pH = 5.6. In the third part 1 μL of 100 mM solution of cysteine was added.

7.8.3 Reaction with mBBr.

Target (10 nmol) and complementary (20 nmol) reactive single-stranded oligonucleotides were mixed (in 10 μL of 0.2 M MOPS pH = 7.4, 0.5 M NaCl) and incubated at 20 °C for 24 h. Then 5 μL of 200 mM stock solution of TCEP (pH = 7.2) was added and reaction was incubated under nitrogen in glove box at RT for 2 hours. Then 0.44 μL of 100 mM stock solution of mBBr was added and mixture was incubated at 25 °C for 30 min. Then reaction was desalting using Nap-5 column to remove any organic compounds and solution was left in the dark at RT. Aliqouts were taken after 1 hour, 21 hours, 69 hours and 92 hours. Final solutions were analysed by RP HPLC “*DNA_Cy3_FLD*” (gradient of 5-10% CH_3CN (0.5 mL/min) in 0.05 M TEAA, pH 7.0) in 10 min. Then CH3CN from 10% to 20% from 10 to 30 min, then from 30% to 80% from 30min to 45 min. UV lamp with wavelength from 240 nm to 800 nm was used as a detector. Fluorescent detector with excitation wavelength at 390 nm and emission at 490 nm was used as additional detector

7.9 Preparing FRET samples

For all experiments, purchased cyanine-labeled DNA strands (ODN_6.2 and ODN_6.4) were stored and used as degased water-stock solutions and stored at -20 ℃. Fluorescence experiments were measured in a quartz cuvette (1 cm) in 0.1 MOPS buffer at pH = 6.9 containing 1 M NaCl at 25℃ without scavengers solutions. Excitation of Cy5-labeled single strand (ODN_6.2) was performed at λ = 650 nm, as well for Cy3-labeled single strand (ODN_6.4) λ = 550 nm.

Excitation for all experiments with pair of labeled complementary DNA strands was at λ = 550 nm. For the fluorescence experiment final concentration of purchased DNA (ODN_6.2 or ODN_6.4) were equal to 2 μM in 100 μL in buffer. To these solution different equivalents (0.1:1, 1:1, 2:1, 10:1 or 20:1) of complementary DNA that was labeled by modular system were added.

7.10 References.

(1) Li, Z.; Zhu, W.; Yu, J.; Ma, X.; Lu, Z.; Xiao, S. *Synth. Commun.* **2006**, *36*, 2613.
(2) Grimm, G. N.; Boutorine, A. S.; Helene, C. *Nucleos. Nucleot. Nucl.* **2000**, *19*, 1943.
(3) Hyrup, B.; Nielsen, P. E. *Biorg. Med. Chem.* **1996**, *4*, 5.
(4) Hartley, J. A.; Bingham, J. P.; Souhami, R. L. *Nucleic Acids Res.* **1992**, *20*, 3175.

Conclusion

In our work we introduced a new strategy for the site-specific modification of long DNA or RNA sequences, respectively. The modular system consists of three building blocks that can be synthesized independently and hence manifest the proposed simple and controllable way for nucleic acid modifications.

As a target for modification, we performed the site-specific transamination of cytosine in a single-step reaction between a nucleophilic reactive group of a reactive sequence and a cytosine containing template DNA (chapter 3) as well as selective alkylation of guanine (chapter 4) or thymidine (chapter 5). We demonstrated, that recognition and duplex formation between reactive and target sequences are the driving forces in our approach. In addition, the exact nature of the reactive group used for modification has a considerable influence on the yield of cross-link formation. For example, in modification of cytosine, the most reactive sequences with primary amino-groups form cross-links with the complementary sequence in low yield, while a hydrazide-containing reactive group significantly improved the yields from 22% to 60%. In the alkylation of thymidine, replacement of the bromoacetate-containing reactive group with the less nucleophilic chloroacetate decreased the yield from 45% to 29%.

Our findings also demonstrate how the specificity of modifications can be altered by changes in the template or the linker architecture. The results clearly reveal that reactive sequences that contain the reactive group in the center flanked by two duplex regions are certainly among the best systems for the mild, fast, selective and site-specific cross-linking of long oligonucleotide sequences. In case of cytosine modification just by using this type of reactive sequence the yield was 2 fold improved.

Additionally, we tested the distance-dependent modification of nucleobases from reactive group and identified that the most reactive target nucleobases are placed at (n+0), (n+1) and (n+2) positions in the templates in all types of modifications.

The modular strategy is also versatile in this respect that a wide variety of other groups and markers e.g., chromophores, fluorophores, enzymes and electron-dense markers, can be attached when using cleavable linkers. This generates a specific reactive group, e.g. an aldehyde or thiol, after cleavage. As an example we demonstrated the possible application of the modular system to label long oligonucleotide templates with different fluorophores. In case of cytosine modification after oxidation of the diol-containing linker we trapped the newly formed aldehyde group with amino-fluoresceine by reductive amination. In another example the disulfide bridge was reductively cleaved and the free thiol group used for selective labeling with monobrominane.

In the last chapter we demonstrated the possibility to use the oligonucleotides modified with our modular system in fluorescence applications. We were able to modify only one of the 14 guanines in a target DNA strand that was placed at position (n+1) relative to reactive group. After cleavage of the cross-linked product, the newly formed thiol group was labeled with maleimido-Cy3 dye. After annealing to a second strand containing a Cy5 dye we observed changes of the emission spectra, i.e. a decrease of donor and an increase of acceptor fluorescence. The results indicate that the site-specific labeling with the modular system shows feasibility in fluorescence applications.